

基于多字典 SBL 算法的宽带声呐角度-速度联合估计

张景铭^{1,2}, 吴 敏^{1,2}, 郝程鹏^{1,2*}, 吴永清^{1,2}

1. 中国科学院声学研究所 水下航行器实验室, 北京 100190

2. 中国科学院大学 电子电气与通信工程学院, 北京 100049

摘要: 针对经典窄带信号处理算法难以被直接应用于宽带信号的问题, 提出了一种适用于宽带 FMCW 声呐的角度-速度联合估计算法。针对宽带信号存在的快-慢时间耦合和空域导向向量随频率改变等问题, 所提算法通过构造多字典稀疏恢复问题实现了在上述问题存在时的角度-速度联合估计, 同时还改善了速度模糊的影响。所提算法使用多字典 SBL 算法求解构造的稀疏恢复问题, 加快了求解速度。数值仿真结果显示所提算法可以在快-慢时间耦合、空域导向向量随频率改变和速度模糊等问题存在的条件下实现角度-速度的高分辨精确估计。

关键词: 调频连续波 (FMCW); 角度-速度联合估计; 稀疏贝叶斯; 宽带信号

文献标志码: A

中图分类号: TN911.72

Joint Angle-Velocity Estimation of Wideband Sonar Based on Multiple Dictionary SBL Algorithm

ZHANG Jingming^{1,2}, WU Min^{1,2}, HAO Chengpeng^{1,2*}, WU Yongqing^{1,2}

1. Institute of Acoustics, Chinese Academy of Sciences, Beijing 100190, China

2. School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing 100049, China

Abstract: To address the challenge that classical narrowband signal processing algorithms cannot be directly applied to wideband signals, an angle-velocity joint estimation algorithm suitable for wideband FMCW sonar is proposed. Addressing issues such as fast time and slow time coupling and frequency-dependent spatial steering vectors inherent in wideband signals, the proposed algorithm constructs a multi-dictionary sparse recovery problem to achieve joint angle-velocity estimation under these conditions while also mitigating the impact of velocity ambiguity. To accelerate the solution process, the proposed algorithm employs a multi-dictionary SBL algorithm to solve the sparse recovery problem. Numerical simulation results demonstrate that the proposed algorithm can achieve high-resolution and accurate angle-velocity estimation under conditions of fast time and slow time coupling, frequency-varying spatial steering vectors, and velocity ambiguity.

Key words: frequency modulated continuous wave (FMCW); joint angle-velocity estimation; sparse Bayesian learning; wideband signals

方位角、距离和速度是运动目标的重要参数, 对它们的估计是阵列信号处理领域的热点问题^[1],

在雷达^[2]、声呐^[3]和无线通信^[4]等领域应用广泛。声呐系统通过估计回波信号的波达角度

基金项目: 国家自然科学基金 (62371446, 62071460, 62001468, 61971412, 62201623, 62471463); 中国科学院青年创新促进会项目 (2023030)。

作者简介: 张景铭 (2000—), 男, 硕士研究生, 研究方向为水声信号处理; 吴敏 (1988—), 女, 博士, 副研究员, 研究方向为水声信号处理、目标探测与成像、新体制声纳信号处理; 郝程鹏 (1975—), 男, 通信作者, 博士, 研究员, 研究方向为水声信号处理、阵列信号处理、信号检测; 吴永清 (1968—), 男, 博士, 研究员, 研究方向为水声定位、水下目标探测和识别。

(direction of arrival, DOA), 多普勒频移和时延来分别对目标的角度、速度和距离进行估计^[5]。相比其他信号, 调频连续波 (frequency modulated continuous wave, FMCW) 信号有更大的时宽-带宽积, 具有测量精度高、平均发射功率低等优势^[6], 因此在声呐系统中应用广泛。通过对回波信号进行解线频调和去斜等处理, 接收信号可被视为多个多维复正弦信号的线性叠加^[7], 其频率由目标回波的角度、目标的速度和距离决定。因此, 对运动目标的参数估计问题可以被转换为多维谱估计问题。

为了突破瑞利限, 以多重信号分类 (multiple signal classification, MUSIC)^[8] 和基于旋转不变技术的信号参数估计 (estimation of signal parameters via rotational invariance techniques, ESPRIT)^[9] 为代表的基于信号子空间的方法被广泛应用于谱估计问题中, 并在之后被推广至高维^[10-12]。

然而此类算法均基于协方差矩阵, 需用充足的样本数以保证性能。与子空间方法相比, 基于压缩感知和稀疏恢复的方法可以在有限样本量的条件下精确估计参数, 甚至在单快拍的情况下也可取得良好效果^[13]。为了避免求解 NP 难的 l_0 范数最小化问题, 基追踪算法等基于 l_1 范数最小化的稀疏算法被广泛应用于运动目标参数估计问题中^[14]。然而, 当问题规模较大时, 基追踪算法的运算复杂度也无法满足应用的需求^[15]。与基追踪算法不同, 稀疏贝叶斯学习 (sparse Bayesian learning, SBL) 算法将具有稀疏特性的系数向量视为参数未知的随机向量, 通过最大化其证据来获得对未知参数的估计^[16]。相比基追踪算法, SBL 算法的运算速度大幅提升, 因此越来越受到关注。

近年来, 为了满足日益增长的高精度距离分辨率的需求, 宽带信号的使用愈加普遍^[17]。随着带宽的增大, 高速目标在一个相干处理间隔 (coherent processing interval, CPI) 中运动的距离可能会超出距离单元, 使得快时间和慢时间的耦合项无法被忽视^[7]。上述耦合项可以用 Keystone 变换和匹配滤波^[18]消除, 但它们需要大量数据以完成内插, 因此无法实现超分辨估计^[17]。另外, 为了实现大带宽, 通常需要更长的脉冲重复间隔, 从而导致速度模糊^[5]。除此之外, 阵列的导向向量随频率的变化也因带宽的增加而无法被忽略, 导致窄带 DOA 估

计方法难以被直接应用。为了解决这一问题, Wax 等人^[19]和 Wang 等人^[20]相继提出了非相干信号子空间方法 (incoherent signal-subspace method, ISSM) 和相干信号子空间方法 (coherent signal-subspace method, CSSM)。然而上述方法均为基于协方差矩阵的, 无法适应小样本的场景。

为了解决上述问题, 本文提出了一种基于多字典 SBL 算法和宽带 FMCW 声呐的角度-速度联合估计算法。通过对信号模型进行推导和分析, 本文将角度-速度的联合估计问题转换为多字典的稀疏恢复问题, 统一解决了宽带信号带来的快-慢时间的耦合问题、速度模糊问题以及空域导向向量随频率变化的问题。为了提升求解速度, 本文使用多字典 SBL 算法实现了对稀疏恢复问题求解。

本文第 1 章对宽带 FMCW 声呐的信号模型进行了简要推导, 并指出了经典窄带处理方法的局限。第 2 章提出了一种全新的基于多字典 SBL 算法的角度-速度联合估计算法。第 3 章展示了数值仿真的结果。最后, 第 4 章对本文内容进行了总结并对未来的研究进行了展望。

1 信号模型

假设共有 N_{Tar} 个待测目标, 其中第 i 个目标距离声呐的径向距离、径向速度和方位角分别为 R_i 、 v_i 和 θ_i , $i \in \{1, 2, \dots, N_{\text{Tar}}\}$ 。发射阵元在一个相干处理间隔中发射一串 Chirp 信号, 其脉宽为 B , 起始频率为 f_c , 脉冲持续时间为 T_c , 脉冲重复间隔为 T_p , 脉冲数为 N_p , CPI 为 $T_{\text{CPI}} = T_p N_p$ 。发射信号 $s(t) = s(\hat{t}, t_m)$ 可表示为:

$$s(\hat{t}, t_m) = \begin{cases} \exp\left(j2\pi\left(f_c t + \frac{1}{2}\mu\hat{t}^2\right)\right), & \hat{t} \in [0, T_c] \\ 0, & \text{其他} \end{cases} \quad (1)$$

其中, $t = t_m + \hat{t}$, $t_m = mT_p$, $m \in \{1, 2, \dots, N_p\}$ 。

发射信号经目标反射后由接收阵接收, 接收阵为包含 N_m 个阵元, 阵元间距为 d 的均匀线阵, 其中 $d = \frac{\lambda_c}{2} = \frac{c}{2f_c}$ 。为了描述接收信号, 首先对接收信号的时延做简要分析。取水中声速 $c=1500$ m/s。令 t 时刻发射的信号经第 i 个目标反射到达接收阵的时延为 $\tau_i(t)$, 其满足:

$$\tau_i(t) = \frac{2(R_i - v_i t)}{v_i + c} \approx \frac{2(R_i - v_i t)}{c} \quad (2)$$

接收阵接收到的对应的反射信号可以表示为：

$$r_i(t) = \alpha_i s(t - \tau_i(t)) \quad (3)$$

其中， $\alpha_i \in \mathbb{C}$ 为发射信号经传播和反射后的损失系数。即

$$\begin{aligned} r_i(\hat{t}, t_m) &= \alpha_i s(t - \tau_i(\hat{t}, t_m)) = \\ &\alpha_i s(\hat{t} - \tau_i(\hat{t}, t_m), t_m) \end{aligned} \quad (4)$$

将式(1)代入式(4)，可得：

$$r_i(\hat{t}, t_m) = \alpha_i \exp(j2\pi\phi_i(\hat{t}, t_m)) \quad (5)$$

其中

$$\begin{aligned} \phi_i(\hat{t}, t_m) &= f_c(t - \tau_i(\hat{t}, t_m)) + \\ &\frac{1}{2}\mu(\hat{t} - \tau_i(\hat{t}, t_m))^2 \end{aligned} \quad (6)$$

由此可得 r_i 的瞬时频率 f_i 为：

$$f_i(\hat{t}, t_m) = \frac{d\phi_i}{dt} \approx f_c + \mu\hat{t} \quad (7)$$

令第 l 个阵元接收到的来自第 i 个目标的反射信号为 $r_{i,l}(t) = r_{i,l}(\hat{t}, t_m)$ 。由于第 i 个目标反射的信号到达第 l 个阵元相对于参考阵元的时延为 $-\frac{dl}{c} \sin(\theta_i)$ ，因此有：

$$\begin{aligned} r_{i,l}(\hat{t}, t_m) &= \alpha_i \exp(j2\pi\phi_i(\hat{t}, t_m)) \cdot \\ &\exp(j2\pi(f_c + \mu\hat{t})\frac{dl}{c} \sin(\theta_i)) \end{aligned} \quad (8)$$

为了实现解线频调脉压，将 $r_{i,l}$ 与参考信号 s 做相关得到差频输出为：

$$\begin{aligned} s_{\text{IF}}^{i,l}(\hat{t}, t_m) &= r_{i,l}(\hat{t}, t_m) \cdot s^*(\hat{t}, t_m) \approx \\ &\alpha_i \exp(j2\pi(f_c + \mu\hat{t})\frac{dl}{c} \sin(\theta_i)) \cdot \\ &\exp\left(-j2\pi\left(\mu\frac{2R_i}{c}\hat{t} + \frac{2v_i f_c}{c}t_m + \mu\frac{2v_i}{c}t_m\hat{t}\right)\right) \end{aligned} \quad (9)$$

当 $\mu T_c dN_m / c \ll 1$ 即 $\frac{BN_m}{2f_c} \ll 1$ 时，式(9)中的

$\mu\hat{t}$ 可以忽略不计，信号可以由窄带模型近似。当 $\mu\frac{2v_i}{c}N_p T_p T_c \ll 1$ 即 $v_i T_{CPI} \ll \frac{c}{2B}$ 时，目标在一个 CPI 中运动的距离不超过一个距离单元，此时式(9)中的 $\mu\frac{2v_i}{c}t_m\hat{t}$ 也可以忽略不计。然而，为了获取更高的距离分辨率，发射信号的带宽 B 也需取

得更宽，使得上述的两项无法被忽略，导致经典的窄带处理方法无法被直接应用。

为了便于后续处理，将式(9)重新整理为如下形式：

$$\begin{aligned} s_{\text{IF}}^{i,l}(\hat{t}, t_m) &= \widetilde{\alpha}_i \exp\left(j2\pi(f_c + \mu\hat{t})\frac{dl}{c} \sin(\theta_i)\right) \cdot \\ &\exp\left(-j2\pi\mu\frac{2R_i}{c}\hat{t}\right) \cdot \exp\left(j2\pi(f_c + \mu\hat{t})\frac{2v_i}{c}t_m\right) \end{aligned} \quad (10)$$

$$\text{其中, } \widetilde{\alpha}_i = \alpha_i \exp\left(-j2\pi f_c \frac{2R_i}{c}\right)。$$

令第 l 个阵元接收到的信号经解线频调脉压后得到的输出为 $y_{\text{IF}}^l(\hat{t}, t_m)$ ，有：

$$y_{\text{IF}}^l(\hat{t}, t_m) = \sum_{i=1}^{N_{\text{Tar}}} s_{\text{IF}}^{i,l}(\hat{t}, t_m) + n_l(\hat{t}, t_m) \quad (11)$$

其中， n_l 为噪声。对上述信号以 f_s 的采样率进行时域采样，并且代入 $t_m = mT_p$ ，采样后的信号 $Y_{\text{IF}}^l(k, m)$ 为：

$$\begin{aligned} Y_{\text{IF}}^l(k, m) &= y_{\text{IF}}^l\left(\frac{k}{f_s}, mT_p\right) = \\ &\sum_{i=1}^{N_{\text{Tar}}} \widetilde{\alpha}_i \exp\left(-j2\pi\mu\frac{2R_i}{cf_s}k\right) \cdot \\ &\exp\left(j2\pi\left(f_c + \frac{\mu}{f_s}k\right)\frac{d \sin(\theta_i)}{c}l\right) \cdot \\ &\exp\left(-j2\pi\left(f_c + \frac{\mu}{f_s}k\right)\frac{2v_i T_p}{c}m\right) + \\ &n_l\left(\frac{k}{f_s}, mT_p\right) \end{aligned} \quad (12)$$

后续将直接对 $Y_{\text{IF}}^l(k, m)$ 进行处理以估计待估参数。

值得一提的是，基于如式(12)的信号模型进行处理还可以改善速度模糊的情况。传统处理方法将快时间和慢时间的耦合忽略不计，得到的速度多普勒项为 $\exp\left(-j2\pi f_c \frac{2v_i T_p}{c}m\right)$ 。此时为了避免速度

模糊，待测的速度应不超过 $v_{\max} = \frac{c}{4f_c T_p}$ 。宽带信号中 T_p 通常取得较大，使得能测出的速度范围非常窄，难以适应实践中对高速目标测速的需求。本文中考虑快时间和慢时间的耦合，得到的速度多普勒项为 $\exp\left(-j2\pi\left(f_c + \frac{\mu}{f_s}k\right)\frac{2v_i T_p}{c}m\right)$ 。为了避免速

度模糊，要求最大速度 v_{\max}^{new} 满足：

$$\text{LCM}\left(\left\{\frac{c}{2T_p\left(f_c + \frac{\mu}{f_s}k\right)}\right\}_{k=1}^{N_t}\right) \geq 2v_{\max}^{\text{new}}$$

其中，LCM 表示最小公倍数， N_t 为快时间采样点数。为了简化讨论，此处考虑 f_c 为 $\frac{\mu}{f_s}$ 的整数倍。

容易证明，此时能测出的最大速度 $v_{\max}^{\text{new}} = \frac{f_s c}{4\mu T_p} > v_{\max}$ ，极大提升了可测速度的范围，改善了速度模糊。

2 基于多字典 SBL 算法的角度-速度联合估计

固定 $Y_{\text{IF}}^l(k, m)$ 的快时间采样指标 k ，令

$$Y_k(l, m) = Y_{\text{IF}}^l(k, m) = \sum_{i=1}^{N_{\text{tar}}} \beta_i \cdot \exp\left(j2\pi\left(f_c + \frac{\mu}{f_s}k\right)\frac{d \sin(\theta_i)}{c}l\right). \quad (13)$$

$$\exp\left(-j2\pi\left(f_c + \frac{\mu}{f_s}k\right)\frac{2v_i T_p}{c}m\right) + n_l\left(\frac{k}{f_s}, mT_p\right)$$

其中， $\beta_{i,k} = \tilde{\alpha}_i \exp\left(-j2\pi\mu\frac{2R_i}{cf_s}k\right)$ 。令 $[Y_k]_{l,m} =$

$$Y_k(l, m), [a_\theta(k, \theta_i)]_l = \exp\left(j2\pi\left(f_c + \frac{\mu}{f_s}k\right)\frac{d \sin(\theta_i)}{c}l\right),$$

$$[a_v(k, v_i)]_m = \exp\left(-j2\pi\left(f_c + \frac{\mu}{f_s}k\right)\frac{2v_i T_p}{c}m\right), \text{ 有：}$$

$$Y_k = \sum_{i=1}^{N_{\text{tar}}} \beta_{i,k} a_\theta(k, \theta_i) a_v(k, v_i)^T + N_k \quad (14)$$

其中， $[N_k]_{l,m} = n_l\left(\frac{k}{f_s}, mT_p\right)$ ， $a_\theta(k, \theta_i) \in \mathbb{C}^{N_m \times 1}$ ，

$a_v(k, v_i) \in \mathbb{C}^{N_p \times 1}$ 。对 Y_k 向量化，可以得到：

$$\begin{aligned} \tilde{y}_k &= \text{vec}(Y_k) = \sum_{i=1}^{N_{\text{tar}}} \beta_{i,k} \tilde{a}(k, (v_i, \theta_i)) + \tilde{n}_k \\ &= A_k \beta_k + \tilde{n}_k \end{aligned} \quad (15)$$

其中， $\tilde{n}_k = \text{vec}(N_k)$ ， $\beta_k = [\beta_{1,k}, \dots, \beta_{N_{\text{tar}},k}]^T$ ，
 $\tilde{a}(k, (v_i, \theta_i)) = a_v(k, v_i) \otimes a_\theta(k, \theta_i)$ ， \otimes 表示 Kronecker 积， $A_k = [\tilde{a}(k, (v_1, \theta_1)), \dots, \tilde{a}(k, (v_{N_{\text{tar}}, \theta_{N_{\text{tar}}}))]$ 。

从式(15)的最后一个等式可以看出，通过将

A_k 扩展为超完备字典，待估参数可由压缩感知算法求得。具体的，首先将 v 和 θ 的取值范围划分为细密的网格，分别构成集合 \mathbb{V} 和集合 Θ 。令

$$\overline{A_v^k} = [\mathbf{a}_v(k, v_1), \mathbf{a}_v(k, v_2), \dots, \mathbf{a}_v(k, v_{|\mathbb{V}|})]$$

$$\overline{A_\theta^k} = [\mathbf{a}_\theta(k, \theta_1), \mathbf{a}_\theta(k, \theta_2), \dots, \mathbf{a}_\theta(k, \theta_{|\Theta|})]$$

其中， $\theta_{i_\theta} \in \Theta, i_\theta = 1, 2, \dots, |\Theta|$ ， $v_{i_v} \in \mathbb{V}, i_v = 1, 2, \dots, |\mathbb{V}|$ ，
 $|\cdot|$ 表示集合的势。将对 A_k 扩展得到的超完备字典用 $\widetilde{A_k}$ 表示，有：

$$\widetilde{A_k} = \overline{A_v^k} \otimes \overline{A_\theta^k} \quad (16)$$

当网格足够细密时， $\exists \widetilde{\beta_k} \in \mathbb{C}^{|\mathbb{V}||\Theta|}$ ，使得：

$$\widetilde{y}_k = \widetilde{A_k} \widetilde{\beta_k} + \widetilde{n}_k \quad (17)$$

此时 $\widetilde{\beta_k}$ 为一稀疏向量，其支撑集可由压缩感知算法求得，并由此得到待估参数 v 和 θ 的估计。令

$$\widetilde{\mathbf{B}} = [\widetilde{\beta_1}, \widetilde{\beta_2}, \dots, \widetilde{\beta_{N_t}}] \quad (18)$$

其中， N_t 为快时间采样的总数。可以看出 $\widetilde{\mathbf{B}}$ 具有组稀疏性，因此可以构造如下优化问题求解

$$\begin{aligned} \min_{\widetilde{\mathbf{B}}} \quad & \|\widetilde{\mathbf{B}}\|_{2,1} \\ \text{s.t.} \quad & \|\widetilde{y}_k - \widetilde{A_k} \widetilde{\beta_k}\|_2 \leq \epsilon_k, k = 1, 2, \dots, N_t \end{aligned} \quad (19)$$

其中， ϵ_k 表示噪声阈值。

上述优化问题虽然可解，但由于问题规模较大而求解复杂。为了满足实时性的需要，下文使用多字典 SBL 方法^[21]实现对参数的估计。

假设 $\widetilde{\beta_k}$ ， \widetilde{n}_k ， $k = 1, 2, \dots, N_t$ 相互独立且均符合 0 均值圆对称高斯分布。由于 $\widetilde{\mathbf{B}}$ 具有组稀疏性，可假设 $\widetilde{\beta_k}$ ， $k = 1, 2, \dots, N_t$ 满足共协方差矩阵先验，即所有的 $\widetilde{\beta_k}$ 的协方差矩阵相同，均为 $\Lambda = \text{diag}(\lambda)$ ，其中 $\lambda = [\lambda_1, \dots, \lambda_{|\mathbb{V}||\Theta|}]^T \in \mathbb{R}^{|\mathbb{V}||\Theta|}$ 。因此有：

$$p(\widetilde{\mathbf{B}}) = \prod_{k=1}^{N_t} p(\widetilde{\beta_k}) = \prod_{k=1}^{N_t} \mathcal{CN}(\widetilde{\beta_k}; \mathbf{0}, \Lambda) \quad (20)$$

假设 $\widetilde{n}_k \sim \mathcal{CN}(\mathbf{0}, \sigma_k^2 I_{N_p N_m})$ ，可求得后验概率为：

$$\begin{aligned} p(\widetilde{y}_1, \widetilde{y}_2, \dots, \widetilde{y}_{N_t} | \widetilde{\mathbf{B}}) &= \prod_{k=1}^{N_t} p(\widetilde{y}_k | \widetilde{\beta_k}) = \\ &\prod_{k=1}^{N_t} \mathcal{CN}(\widetilde{y}_k; \widetilde{A_k} \widetilde{\beta_k}, \sigma_k^2 I_{N_p N_m}) \end{aligned} \quad (21)$$

结合式(20)和式(21)可得：

$$p(\tilde{\mathbf{y}}_1, \tilde{\mathbf{y}}_2, \dots, \tilde{\mathbf{y}}_{N_t}) = \prod_{k=1}^{N_t} p(\tilde{\mathbf{y}}_k) = \prod_{k=1}^{N_t} p(\tilde{\mathbf{y}}_k | \tilde{\boldsymbol{\beta}}_k) p(\tilde{\boldsymbol{\beta}}_k) d\tilde{\boldsymbol{\beta}}_k = \prod_{k=1}^{N_t} \mathcal{CN}\left(\tilde{\mathbf{y}}_k; \mathbf{0}, \widetilde{\mathbf{A}}_k \mathbf{A} \widetilde{\mathbf{A}}_k^H + \sigma_k^2 \mathbf{I}_{N_p N_m}\right)$$

对式(22)取对数可得:

$$\log p(\tilde{\mathbf{y}}_1, \tilde{\mathbf{y}}_2, \dots, \tilde{\mathbf{y}}_{N_t}) = -\sum_{k=1}^{N_t} \left(\log \left(\det \left(\widetilde{\mathbf{A}}_k \mathbf{A} \widetilde{\mathbf{A}}_k^H \right) \right) + \tilde{\mathbf{y}}_k^H \left(\widetilde{\mathbf{A}}_k \mathbf{A} \widetilde{\mathbf{A}}_k^H \right)^{-1} \tilde{\mathbf{y}}_k \right) + const \right) \quad (23)$$

通过最大化上述对数概率密度即可获得对 λ 的估计 $\hat{\lambda}$, 即

$$\begin{aligned} \hat{\lambda} &= \arg \max_{\lambda} \log p(\tilde{\mathbf{y}}_1, \tilde{\mathbf{y}}_2, \dots, \tilde{\mathbf{y}}_{N_t}) = \\ &= \arg \min_{\lambda} \sum_{k=1}^{N_t} \left(\log \left(\det \left(\widetilde{\mathbf{A}}_k \mathbf{A} \widetilde{\mathbf{A}}_k^H \right) \right) + \tilde{\mathbf{y}}_k^H \left(\widetilde{\mathbf{A}}_k \mathbf{A} \widetilde{\mathbf{A}}_k^H \right)^{-1} \tilde{\mathbf{y}}_k \right) \right) \quad (24) \end{aligned}$$

式(24)可通过Nannuru等人^[21]提出的迭代算法求解, 具体的,

$$\lambda_i^{\text{new}} = \lambda_i^{\text{old}} \left(\frac{\sum_{k=1}^{N_t} \text{tr} \left(\Sigma_{\tilde{\mathbf{y}}_k}^{-1} [\widetilde{\mathbf{A}}_k]_{:,i} [\widetilde{\mathbf{A}}_k]^H \Sigma_{\tilde{\mathbf{y}}_k}^{-1} \mathbf{S}_{\tilde{\mathbf{y}}_k} \right)}{\sum_{k=1}^{N_t} \text{tr} \left(\Sigma_{\tilde{\mathbf{y}}_k}^{-1} [\widetilde{\mathbf{A}}_k]_{:,i} [\widetilde{\mathbf{A}}_k]^H \right)} \right)^b \quad (25)$$

其中, $\Sigma_{\tilde{\mathbf{y}}_k} = \widetilde{\mathbf{A}}_k \mathbf{A} \widetilde{\mathbf{A}}_k^H + \sigma_k^2 \mathbf{I}_{N_p N_m}$, $\mathbf{S}_{\tilde{\mathbf{y}}_k} = \tilde{\mathbf{y}}_k \tilde{\mathbf{y}}_k^H$, b 为控制收敛速率的可调参数。式(25)中需要用到噪声功率 σ_k^2 , 当其未知时, 可以对其做如下估计:

$$\widehat{\sigma_k^2} = \frac{1}{N_m N_p - N_{\text{Tar}}} \text{tr} \left(\left(\mathbf{I}_{N_p N_m} - [\widetilde{\mathbf{A}}_k]_{:,T} [\widetilde{\mathbf{A}}_k]_{:,T}^\dagger \right) \mathbf{S}_{\tilde{\mathbf{y}}_k} \right) \quad (26)$$

其中, T 为 λ 的支撑集, 即 λ 中最大的 N_{Tar} 个元素对应的索引, $(\cdot)^\dagger$ 表示伪逆。

最后, 方位角 θ 和速度 v 的联合谱 $Spec$ 可由估计得到的 $\hat{\lambda}$ 重新排列为 $|\Theta| \times |\mathbb{V}|$ 的矩阵得到, 即

$$Spec = \text{reshape}(\hat{\lambda}, |\Theta|, |\mathbb{V}|) \quad (27)$$

3 仿真结果

令目标个数 $N_{\text{Tar}} = 3$, 目标离声呐的距离分别

为4 m、19 m和10 m, 方位角分别为 11° 、 12° 和 13° , 径向速度分别为10 m/s、-10 m/s和9 m/s。设置发射信号起始频率 f_c 为1500 Hz, 带宽为1500 Hz, 脉冲持续时间为0.1 s, 脉冲重复间隔为0.5 s, 脉冲数为5, 接收阵列为一均匀线阵, 其阵元数为5, 阵元间距为 $d = c / (2f_c)$ 。可以看出, 若忽略快时间和慢时间的耦合, 则此时可以估计的最大速度为 $v_{\text{max}} = 0.5$ m/s, 会出现速度模糊。考虑耦合项后, 根据第2章的讨论, 可以估计的最大速度为 $v_{\text{max}} = 20$ m/s, 此时不存在速度模糊。当不考虑噪声时, 分别使用窄带MUSIC算法、ISSM算法和所提基于多字典SBL算法得到的角度-速度联合估计结果如图1所示。

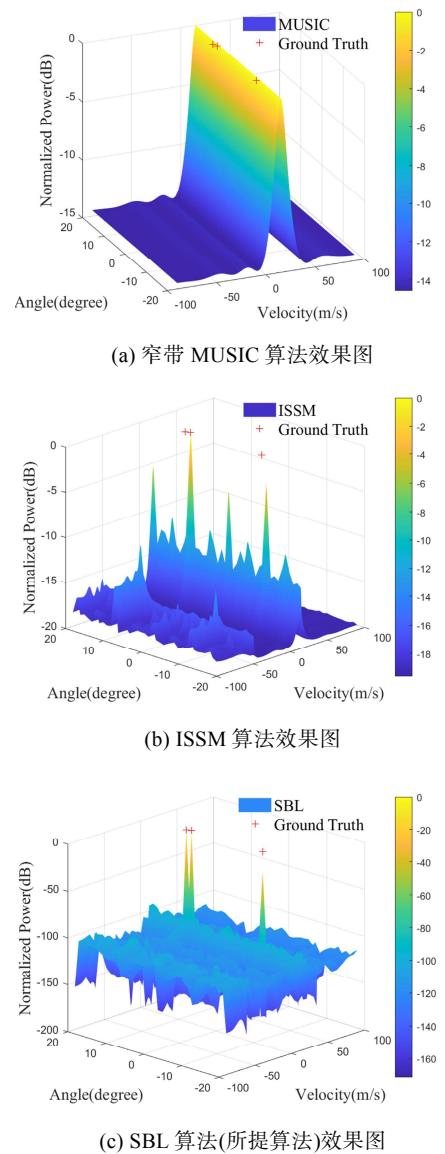


图1 无噪声情况下角度-速度联合估计效果

从图 1(a)可看出,窄带 MUSIC 算法完全无法分辨信号,由此可看出式(9)中的耦合项不能被忽略。图 1(b)中宽带 ISSM 算法由于快拍数量不足,导致估计结果存在大量伪峰,且在两个目标参数相近时无法将其分辨。相比之下,图 1(c)中本文所提算法可以实现目标方位角和速度的准确估计,且峰值最尖锐,旁瓣最低,分辨率最高。

为验证所提算法噪声鲁棒性,图 2 给出了不同信噪比 (signal-to-noise ratio, SNR) 下角度、距离的估计精度。添加高斯白噪声,当信噪比为 0~50 dB 时,进行 100 次蒙特卡洛仿真,角度和速度的估计的根均方误差 (root mean square error, RMSE) 如图 2 所示。

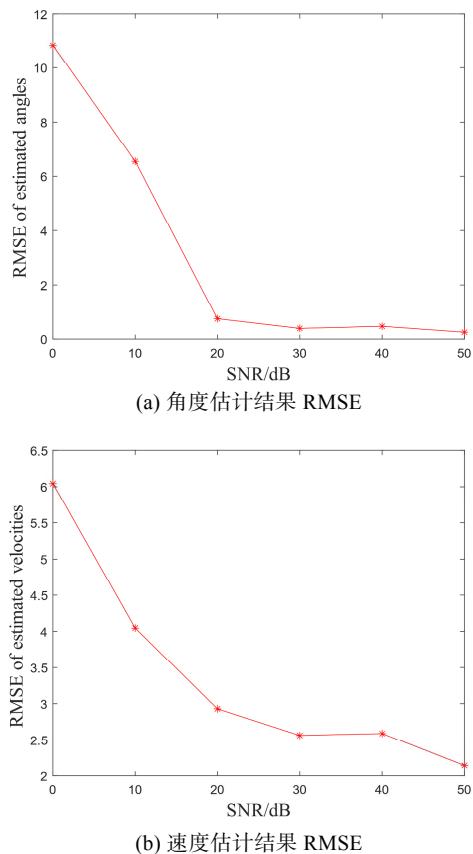


图 2 角度和速度估计 RMSE 与 SNR 关系图

从图 2(a)和图 2(b)中可以看出,当信噪比高于 20 dB 时,所提方法对角度和速度的估计均可取得较高的精确度,展现了所提算法的鲁棒性。

4 结束语

本文提出了一种针对宽带 FMCW 声呐的角度-速度联合估计算法。通过对宽带 FMCW 声呐信号

进行建模,推导了由宽带信号导致的快-慢时间耦合以及空域导向向量随频率变化等问题在信号模型上的体现。基于推导的信号模型,本文将角度-速度联合估计问题转化为多字典稀疏恢复问题,并使用多字典 SBL 算法对其求解以加快运算速度。理论和仿真实验证明所提方法还可对速度模糊问题进行改善。通过开展数值仿真验证,所提方法可在带宽与起始频率相当以及目标在一个 CPI 内跨越距离单元的情况下对目标的方位角和速度实现高分辨的精确估计。然而,本文目前使用的估计方法是基于网格的,容易受到网格失配的影响,未来研究将进一步探索无网格的估计方法,进一步提升估计的精确度。

参考文献:

- [1] ZHOU Z, DAI W. Gridless joint delay-DoA-Doppler estimation using OFDM signals: a multilevel hankel matrix approach[C]//2022 30th European Signal Processing Conference (EUSIPCO). IEEE, 2022: 1821-1825.
- [2] YU F, GUO S, HAO X, et al. AT-BLR: AOA and TD based multimaterial building layout reconstruction[J]. IEEE Transactions on Geoscience and Remote Sensing, 2024.
- [3] YANG J, YANG Y, LEI B. An efficient compressed sensing-based DOA estimation method in nested MIMO sonar[C]//OCEANS 2017-Aberdeen. IEEE, 2017: 1-4.
- [4] ZHAI W, WANG X, WANG X, et al. Optimal sparse MIMO transceiver design for joint automotive sensing and communications[C]//2023 IEEE International Conference on Acoustics, Speech, and Signal Processing Workshops (ICASSPW). IEEE, 2023: 1-5.
- [5] 田坦.声呐技术[M]. 哈尔滨: 哈尔滨工程大学出版社, 2010.
- [6] LI Y, LONG Q, WU Z, et al. Low-complexity joint 3D super-resolution estimation of range velocity and angle of multi-targets based on FMCW radar[J]. Sensors, 2022, 22(17): 6474.
- [7] 保铮, 邢孟道, 王彤. 雷达成像技术[M]. 北京: 电子工业出版社, 2005.
- [8] SCHMIDT R O. A signal subspace approach to multiple emitter location and spectral estimation[M]. Stanford University, 1982.
- [9] PAULRAJ A, ROY R, KAILATH T. A subspace rotation approach to signal parameter estimation[J]. Proceedings of the IEEE, 1986, 74(7): 1044-1046.

(下转第 278 页)

基于球面波分解的 ArcSAR 三维成像算法

张 远，乔一雪，王彦平，王伯伦

北方工业大学 信息学院 人工智能系，北京 100144

摘要：圆周合成孔径雷达（ArcSAR）三维成像技术是一种前沿的遥感观测方法，通过结合合成孔径雷达高分辨率成像和圆周扫描技术的全方位观察特性，实现对目标区域的立体、全向观察。现有的地基合成孔径雷达（GBSAR）成像技术存在系统复杂、计算量大以及数据采集过程复杂等问题。为克服传统算法的局限性，在二维平面 ArcSAR 成像算法的基础上，提出基于球面波分解的 ArcSAR 三维成像算法。该算法在保持 ArcSAR 技术全方位观察特性的同时，增加了高度向分辨率，使得成像结果更为清晰。算法的技术核心是通过将球面波分解为平面波的积分形式，融合极坐标格式算法，实现单次摆扫条件下，获取目标区域的全景、全向立体影像，突破原始极坐标格式算法所存在的聚焦问题和近似问题。实验结果表明，该算法可以对人体模型进行清晰的三维摆扫成像，且在成像速度和受地面干扰程度等方面均表现良好。

关键词：地基合成孔径雷达（GBSAR）；圆周合成孔径雷达成像（ArcSAR 成像）；三维成像

文献标志码：A

中图分类号：TP751.1

ArcSAR 3D Imaging Algorithm Based on Spherical Wave Decomposition

ZHANG Yuan, QIAO Yixue, WANG Yanping, WANG Bolun

Department of Artificial Intelligence, School of Information Science and Technology, North China University of Technology, Beijing 100144, China

Abstract: Arc synthetic aperture radar (ArcSAR) three-dimensional imaging technology is a cutting-edge remote sensing observation method, by combining the high resolution imaging of synthetic aperture radar and the all-round observation characteristics of circular scanning technology, to realize the three-dimensional and omnidirectional observation of the target area. The existing ground-based synthetic aperture radar (GBSAR) imaging technology has some problems, such as complex system, large amount of calculation and complicated data acquisition process. In order to overcome the limitations of traditional algorithms, a three-dimensional imaging algorithm of ArcSAR based on spherical wave decomposition is proposed on the basis of two-dimensional planar ArcSAR imaging algorithm. This algorithm not only maintains the all-round observation characteristics of ArcSAR technology, but also increases the altitude resolution, which makes the imaging results clearer. The technical core of the algorithm is to decompose the spherical wave into the integral form of the plane wave and integrate the polar coordinate format algorithm to obtain the panoramic and omnidirectional stereo image of the target area under the condition of a single swing scan, which breaks through the focusing problem and approximation problem existing in the original polar coordinate format algorithm. The experimental results show that the proposed algorithm can carry out clear three-dimensional swing imaging of human body model, and the imaging speed and the degree of ground interference are good.

Key words: ground-based synthetic aperture radar (GBSAR); arc synthetic aperture radar imaging (ArcSAR imaging); three-dimensional imaging (3D imaging)

作者简介：张远（1983—），男，博士，副教授，研究方向为信号处理、毫米波雷达；乔一雪（1999—），女，硕士研究生，研究方向为雷达信号处理、合成孔径雷达成像；王彦平（1976—），男，博士，教授，研究方向为智能雷达、雷达形变监测、智能预警技术；王伯伦（2000—），男，硕士研究生，研究方向为基于毫米波雷达点云的人体行为识别。

近年来，在地面对目标监测场景下，地基合成孔径雷达（ground-based synthetic aperture radar，GBSAR）因其卓越的高精度监测特性在遥感领域备受瞩目。GBSAR 系统具有非接触、全天时、全天候的工作能力，这一特性使得其在遥感监测中不依赖于外部光照条件，且几乎不受气候因素的制约。特别在地形地貌复杂多变、自然环境恶劣的地区，GBSAR 系统凭借其独特的技术优势，能够稳定且有效地进行数据采集和地物识别^[1-2]。相较于传统的合成孔径雷达（synthetic aperture radar，SAR）系统，GBSAR 系统在成像方式上更加灵活，可以通过调整雷达参数和观测角度，获得不同分辨率和覆盖范围的图像，满足多样化的应用需求。基于以上优点，GBSAR 成为国内外各监测领域研究的热点，且常用于露天矿边坡^[3-4]的稳定性监测、大坝安全监测^[5]、滑坡救援^[6-7]。

GBSAR 系统主要分为直线 SAR 和圆周 SAR 两种类型^[8-10]。直线 SAR 系统通过沿直线轨道移动雷达设备来形成合成孔径，从而实现对目标场景的二维成像。Röedelsperger 等人提出了 IBIS-L 地基 SAR 系统^[11]，用于采石场的位移监测以及活火山的“火山快速反应系统”。MetaSensing 的 FastGBSAR 系统^[12]能够在真实孔径雷达模式下精确测量位移，适用于桥梁等结构振动监测，还能够生成二维位移图像，适用于水坝、山体滑坡等区域监测。然而，直线 SAR 系统也存在着视野范围有限^[13]、受地形起伏影响大等缺点，这些缺点限制了该系统在特定的大视场场景下的应用效果。为了改善直线 SAR 系统的视野限制，学者们提出了圆周 SAR 的实现形式。圆周 SAR 系统通过雷达设备在圆周轨道上的运动，采集不同角度的雷辔回波信号进行成像，实现对周围环境的全方位监测，该系统具有高角度分辨率、视野广阔等优点，相较于直线 SAR 更适用于桥梁监测和建筑物监测的场景。IDS 公司开发的 IBIS-ArcSAR 系统^[14]能实现 360° 水平覆盖，且具有恒定的角度分辨率，通过结合多输入多输出（multiple-input multiple-output，MIMO）雷达通道，能够重建矿山的垂直高度以及数字高程模型，增强二维 SAR 图像的聚焦，克服露天矿山监测中的方位角覆盖限制。中科院电子研究所的 Arc FMCW SAR 系统^[15]通过天线旋转形成合成孔

径，具有扫描大方位角范围的优势，实验验证该系统适用于滑坡探测和数字高程模型提取。但圆周 SAR 系统也存在着结构复杂和数据处理难度较大的缺点，从而带来了更高的技术挑战和成本。为进一步提升 GBSAR 系统的性能和应用范围，克服直线 SAR 和圆周 SAR 的局限性，更大视野的 ArcSAR 三维成像系统应运而生。ArcSAR 系统结合了直线 SAR 和圆周 SAR 的优点，既保证了较广的视野，又在一定程度上简化了系统结构，通过优化雷达设备的运动轨迹和数据处理算法，实现了更大范围的监测和更高精度的成像，致使 ArcSAR 系统不仅具有广阔的视野和灵活的监测能力，还能够在复杂地形和恶劣环境下保持稳定的性能。

本文提出一种基于球面波分解的 ArcSAR 三维成像算法，在传统二维平面 ArcSAR 成像算法的基础上增加了高度分辨率，本文算法相较于传统 ArcSAR 算法更适用于大视场场景，且运算速度较快，受地面干扰更少，成像结果更清晰。本算法对接收到的目标物体的球面波形式回波信号进行分解处理，从而得到其平面波表示，根据 ArcSAR 系统的扫描角度和方向构建滤波器，通过快速傅里叶变换（fast Fourier transform，FFT）进而有效消除平面波积分形式下的高阶相位误差，提高成像的精度和稳定性。为进一步分离角度方向和距离方向的参数，准确掌握信号特性，对信号进行 Stolt 插值处理。将插值处理后的信号进行二维快速傅里叶逆变换（inverse fast Fourier transform，IFFT），实现信号由频率域到空间域的转换，得到其二维点目标聚焦结果，再次进行高度向 IFFT 最后得到目标的三维成像结果。

根据基于球面波分解的 ArcSAR 三维成像算法的研究内容，本文组织结构如下：第 1 章介绍本文三维成像系统的几何模型与信号模型；第 2 章对本文三维成像算法进行理论推导，第 3 章根据本文算法进行仿真实验，证明其可行性，第 4 章利用本文算法进行真实数据实验，分析本文算法的成像质量，第 5 章对全文进行总结。

1 ArcSAR 三维成像系统信号模型

ArcSAR 三维成像系统的几何模型如图 1 所示。ArcSAR 三维成像系统由线性阵列雷达和圆周

转台组合而成，雷达选用线性调频连续波，通过一维线性阵列雷达顺时针旋转，等效为合成孔径雷达进行扫描，由此得到扫描目标的三维球面波形式回波信号。雷达 S 的旋转半径为 r ，设雷达 S 高度的平面为 z_s ， θ_s 是雷达 S 在 XOY 平面投影 S' 的旋转角， θ_p 是目标 F 在 XOY 平面投影 P 的旋转角。目标 F 的坐标为 (x, y, z) ，雷达 S 的坐标为 $(r \cos \theta_s, r \sin \theta_s, z_s)$ 。

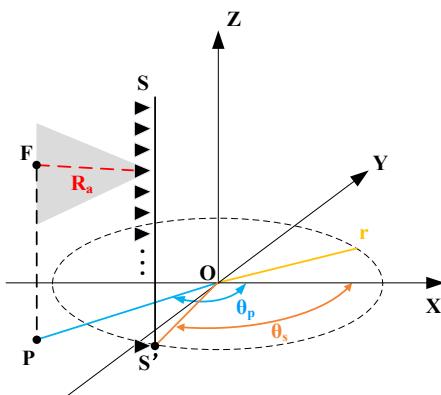


图 1 ArcSAR 三维成像系统的几何模型

则目标 F 到雷达 S 的目标斜距 R_a 表达式为：

$$R_a = \sqrt{r^2 + r_{po}^2 - 2r \cdot r_{po} \cos(\theta_s - \theta_p)} \quad (1)$$

收到的回波信号表达式为：

$$S(\theta_s, k, z_s) = \exp\{-j2\pi k R_a\} \quad (2)$$

为进一步研究信号时频特性，引入时频替换

$$f_r = \text{chirp_rate} \cdot \tau, \tau \text{ 代表脉冲宽度}, \tau \in \left[-\frac{T_p}{2}, \frac{T_p}{2}\right],$$

chirp_rate 表示信号调频率。则上式中波数 k 可以表示为：

$$k = \frac{f_0 + f_r}{c/2} \quad (3)$$

式中， f_0 为发射信号的载频。

以上是 ArcSAR 三维成像系统算法模型的建立，以及回波信号的通用处理方法。

2 基于球面波分解的 ArcSAR 三维成像算法

本章将针对基于球面波分解的 ArcSAR 三维成像算法进行推导。首先，对回波信号 $S(\theta_s, k, z_s)$ 进行高度向的 FFT，把回波信号从空间域转换到波数域。此时的回波信号为：

$$S(\theta_s, k, k_z) = \exp\left\{j\left(-2\pi\sqrt{k^2 - k_z^2} \cdot R_a\right)\right\} \cdot \exp\{-j2\pi k_z \cdot z\} \quad (4)$$

式中， k_z 为波数在高度向的分量。

为精确估计未知值，对回波信号沿高度向进行 Stolt 映射，一般采用插值法实现估算，结果如下：

$$\sqrt{k^2 - k_z^2} \rightarrow k \quad (5)$$

将式 (5) 带入到式 (4) 中，得到插值处理后的信号表达式：

$$S(\theta_s, k, k_z) = \exp\{-j2\pi k \cdot R_a\} \cdot \exp\{-j2\pi k_z \cdot z\} \quad (6)$$

式 (6) 中所得相位为球面波形式的目标相位，由文献[15]对其进行分解，得到平面波积分形式下的表达式：

$$\begin{aligned} \frac{1}{-j2\pi k R_a} \exp\{-j2\pi k R_a\} = \\ \oint \exp\{j2\pi k [(r \cos \theta_s - x) \cos \varphi \sin \eta + (r \sin \theta_s - y) \sin \varphi \sin \eta + 0 \cdot \cos \eta]\} \sin \eta d\eta d\varphi \end{aligned} \quad (7)$$

当球面波被分解后，其成像平面呈现出二维特性，即 $\eta = \frac{\pi}{2}$ ，由此可得出 ArcSAR 三维成像算法在进行球面波分解时，只涉及一重积分形式，故将 $\eta = \frac{\pi}{2}$ 带入到上式：

$$\begin{aligned} \frac{1}{-j2\pi k R_a} \exp\{-j2\pi k R_a\} = \\ \oint \exp\{j2\pi k [(r \cos \theta_s - x) \cos \varphi + (r \sin \theta_s - y) \sin \varphi]\} d\varphi \end{aligned} \quad (8)$$

ArcSAR 三维成像系统的球面波分解如图 2 所示，依据球面波分解公式，将式 (8) 进行重构。

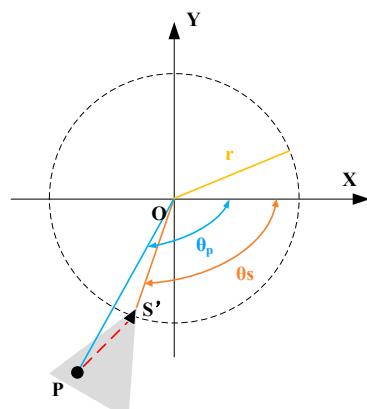


图 2 ArcSAR 三维成像系统的球面波分解示意图

$$\begin{aligned} \exp\{-j2\pi kR_a\} &= \\ -j2\pi kR_a \oint_0^{2\pi} \exp\{j2\pi k[(r\cos\theta_s - x)\cos\varphi + (r\sin\theta_s - y)\sin\varphi]\} d\varphi &= \\ j2\pi R_a \oint_0^{2\pi} \exp\{-j2\pi(k_x \cdot x + k_y \cdot y)\} d\varphi &= \\ \exp\{j2\pi k \cdot r \cos(\varphi - \theta)\} dk_z d\varphi & \end{aligned} \quad (9)$$

式中,

$$k_x = k \cos\varphi, k_y = k \sin\varphi \quad (10)$$

式(9)中,当对信号幅度的因素选择忽视时,等式右侧可视为两部分进行卷积,

$$\exp\{-j2\pi kR_a\} = f(\theta, k) * g(\theta, k) \quad (11)$$

式中,

$$f(\theta, k) = \exp\{j2\pi k \cdot r \cdot \cos\theta\} \quad (12)$$

$$g(\theta, k) = \exp\left\{j[-2\pi k_x \cdot x - 2\pi k_y \cdot y]\right\}_{\varphi=\theta} \quad (13)$$

通过对式(11)等式两边沿角度方向同时FFT,由卷积定理得出 $g(\theta, k)$,可得:

$$\begin{aligned} \mathcal{F}_\theta\{SS(\theta, K)\} &= \mathcal{F}_\theta\{f(\theta, K)\} * g(\theta, K) = \\ F(\xi, K) \cdot G(\xi, K) \end{aligned} \quad (14)$$

其中, ξ 是角度方向的频率。可得:

$$G(\xi, K) = \frac{\mathcal{F}_\theta\{SS(\theta, K)\}}{F(\xi, K)} \quad (15)$$

对式(15)等式两侧同时IFFT,可得:

$$g(\theta, K) = \mathcal{F}_\xi^{-1}\left\{\frac{\mathcal{F}_\theta\{SS(\theta, K)\}}{F(\xi, K)}\right\} \quad (16)$$

对回波数据沿角度方向进行FFT可以求解出上式中的 $\mathcal{F}_\theta\{SS(\theta, K)\}$ 。对式(12)沿角度方向进行FFT可以求解出上式中的 $F(\xi, K)$ 。经处理后,得到方向滤波后的 $g(\theta, K)$,即

$$g(\theta, k) = \exp\left\{j[-2\pi k_x \cdot x - 2\pi k_y \cdot y]\right\} \quad (17)$$

式中,

$$k_x = k \cos\varphi, k_y = k \sin\varphi \quad (18)$$

由于 $g(\theta, K)$ 是极坐标系形式下的表达式,为方便后续处理,需采用极坐标转直角坐标插值的方法对其进行解耦,处理过程如图3。其中, θ_1 表示开始扫描角度, θ_2 表示结束扫描角度; K_1 代表最小波数, K_2 代表最大波数。由图可得,系统参数设定后,数据区域面积固定不变,算法的计算量不受实际场景影响,因此本文算法在计算效率较高。

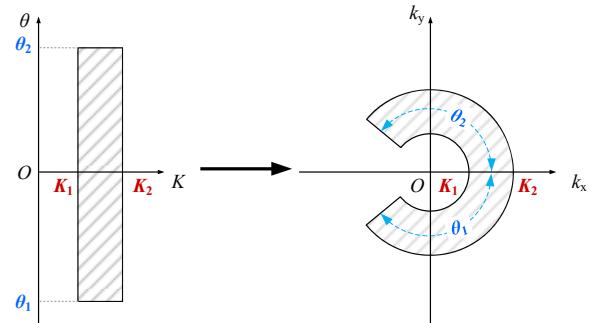


图3 插值示意图

经插值处理后的 $g(\theta, K)$ 表示为:

$$g_{xy}(k_x, k_y) = \exp\left\{j[-2\pi k_x \cdot x - 2\pi k_y \cdot y]\right\} \quad (19)$$

为得到二维点目标聚焦结果,对式(19)分别沿 k_x 和 k_y 方向进行IFFT,再沿高度向做IFFT,得到最终三维聚焦结果。

本文算法处理流程如图4所示。

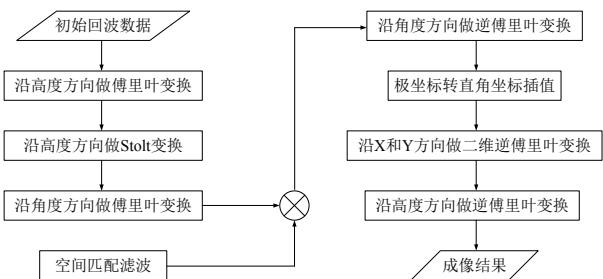


图4 算法处理流程图

3 仿真实验与分析

本文以Matlab 2021a软件为基础,为保证对人体模型的穿透效果,实验依托26.1 GHz雷达板作为硬件支持,通过建立点目标进行仿真实验,对比仿真实验结果与理想模型中点目标的位置,对本文中所介绍的算法进行仿真验证。仿真实验全程严格按照机械扫描设备的各项参数进行设置和执行,其参数如表1所示。

表1 雷达仿真参数

参数	参数值
起始频率/GHz	24.4
终止频率/GHz	27.8
信号带宽/GHz	3.4
扫描角度/(°)	360
方位向波数角宽度/(°)	60
高度向波数角宽度/(°)	90
角度采样间隔/(°)	0.3
高度/m	0.2

本实验中点目标严格按照 $0.2\text{ m} \times 0.2\text{ m}$ 的间隔标准进行排列, 如图 5 所示, 点目标在平面中呈现出清晰、有规划分布。

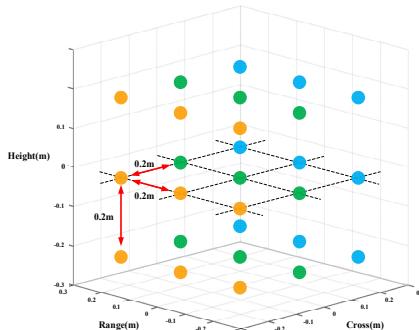


图 5 点目标仿真点阵图

本算法的仿真结果如图 6 所示, 通过将图 6 中 Height-Range、Cross-Range、Height-Cross 的不同方位成像结果与图 5 中三维点目标的点阵图进行对比, 可以明显看出仿真实验结果中点目标呈现

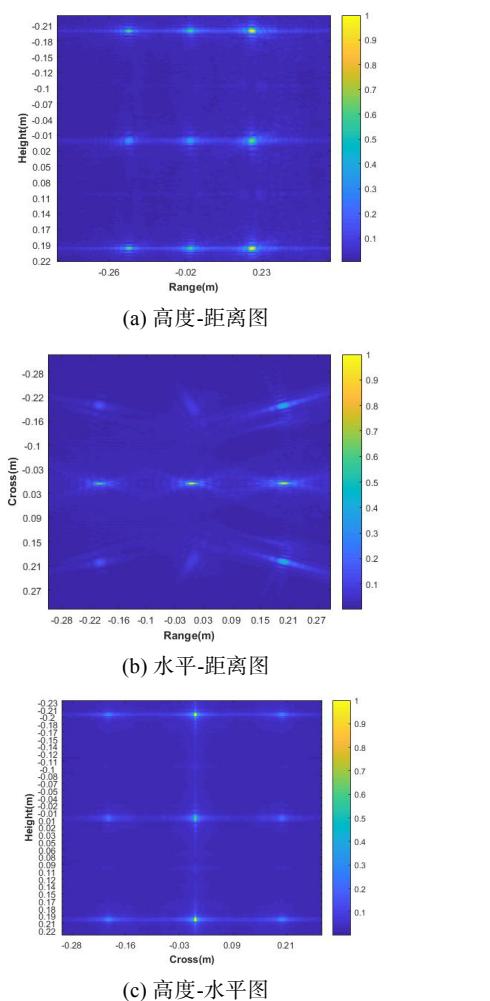


图 6 本文算法仿真结果图

$0.2\text{ m} \times 0.2\text{ m}$ 均匀分布, 探测出的目标位置、数量、形状均与图 5 中建立的模型一致, 点目标显示较清晰, 只存在微小的模糊、叠掩现象, 表明本文算法具有高分辨率且结果准确度较高。

4 真实数据实验

本研究在 ArcSAR 三维成像系统下, 进行真实数据实验, 依托载频为 26.1 GHz 的雷达芯片, 雷达信号带宽为 3.4 GHz, 具备高分辨率精确测距能力。其实验场景如图 7 所示, 以确保覆盖目标区域, 雷达旋转半径设定为 0.4 m, 系统方位向和高度向波束宽度依次为 60° 和 90° 。



图 7 基于 ArcSAR 三维成像系统扫描人体模型图

实际实验数据来自该系统在人体模型上的扫描实验, 人体模型光学图像如图 8 所示, 人体模型放置在距离雷达 0.6 m 的位置, 通过系统面对人体模型顺时针旋转进行数据的采集, 使用本文算法对采集到的数据进行成像处理, 结果如图 9 所示。

通过图 9 展示的成像结果可以看出本文算法的分辨率较高, 能够更加快速、精确地识别出人体模型的基本轮廓。



图 8 人体模型光学图像

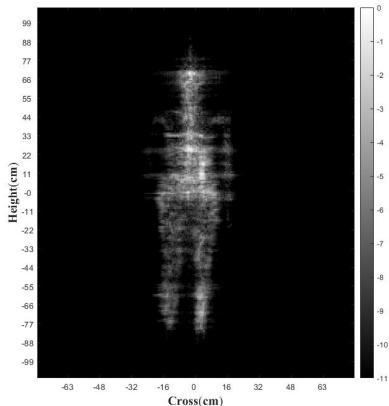


图9 基于球面波分解的ArcSAR三维成像
算法扫描人体模型结果

5 结束语

本文从成像模型、算法推导、仿真实验验证以及实际数据实验四个方面介绍了用于 ArcSAR 三维成像系统的基于球面波分解的 ArcSAR 三维成像算法。算法首先通过沿高度方向执行 FFT，将接收到的回波信号从时域转换至波数域，分析信号在频谱空间中的特性。采用 Stolt 插值技术，沿高度方向对变换后的回波信号进行处理，进而获得球面波形式的目标相位，即目标在三维空间中的位置信息。为简化信号处理过程，对信号相位进行相位展开操作，以转化为平面波的积分形式。在此基础之上，为精确提取出目标信息，同时消除角度向高阶误差，构造空间匹配滤波器。通过插值处理，实现角度方向与距离方向的解耦，将信号从极坐标系统映射到直角坐标系统。最终沿高度方向执行 IFFT，将处理后的信号从波数域转换回时域。对该算法进行仿真模拟实验，验证本算法的成像性能，对本算法进行真实数据实验，进一步体现出算法的可行性。接下来将围绕本文算法展开更多仿真实验和数据实验的对比研究工作，针对户外场景进一步提升算法能力，实现算法的工程化，并对算法性能进行提升。

参考文献：

- [1] 林赟, 宋扬, 王彦平, 等. 旋转扫描地基 SAR 大视场快速成像算法[J]. 信号处理, 2019, 35(3): 499-506.
- [2] 赵子恒, 林赟, 王彦平, 等. 圆周扫描地基 SAR 地形表面提取辅助的三维成像算法[J]. 信号处理, 2022, 38(1): 83-91.
- [3] 王彦平, 黄增树, 谭维贤, 等. 地基 SAR 干涉相位滤波优化方法[J]. 信号处理, 2015, 31(11): 1504-1509.
- [4] 王彦平, 崔紫维, 曹琨, 等. 基于注意力网络的地基 SAR 时序差分相位分类方法[J]. 信号处理, 2021, 37(7): 1207-1216.
- [5] 刘龙龙, 张继贤, 王世杰, 等. 新型 FMCW 地基 SAR 和三维激光扫描仪在大坝变形监测中的应用[J]. 测绘通报, 2019(3): 76-80.
- [6] FERRIGNO F, GIGLI G, FANTI R, et al. GB-InSAR monitoring and observational method for landslide emergency management: the Montaguto earthflow (AV, Italy)[J]. Natural Hazards and Earth System Sciences, 2017, 17(6): 845-860.
- [7] WANG Y P, LV S, BAI Z C, et al. Research on GBSAR deformation extraction method based on three-threshold [C]//2018 11th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), Beijing, China, 2018:1-6.
- [8] 张健丰, 付耀文, 张文鹏, 等. 圆迹合成孔径雷达成像技术综述[J]. 系统工程与电子技术, 2020, 42(12): 2716-2734.
- [9] 申文杰, 林赟, 王彦平, 等. 基于子孔径图像的单通道圆迹 SAR 动目标轨迹反演算法研究[J]. 信号处理, 2017, 33 (3A):41-46.
- [10] WANG Y P, ZHANG Q M, LIN Y, et al. Multi-phase-center sidelobe suppression method for circular GBSAR based on sparse spectrum[J]. IEEE Access, 2020, 8: 133802-133816.
- [11] ROEDELSPERGER S, LAEUFER G, GERSTENECKER C, et al. Monitoring of displacements with ground-based microwave interferometry: IBIS-S and IBIS-L[J]. Journal of Applied Geodesy, 2010, 4(1): 41-54.
- [12] NOTARNICOLA C, PALOSCIA S, PIERDICCA N, et al. MetaSensing's FastGBSAR: ground based radar for deformation monitoring[C]//Proceedings of SPIE, 2014, 9243: 924318.
- [13] ZHANG Y, ZHANG Q M, WANG Y P, et al. An approach to wide-field imaging of linear rail ground-based SAR in high squint multi-angle mode[J]. Journal of Systems Engineering and Electronics, 2020, 31(4): 722-733.
- [14] VIVIANI F, MICHELINI A, MAYER L, et al. IBIS-ArcSAR: an innovative ground-based SAR system for slope monitoring[C]//IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain. IEEE, 2018: 1348-1351.
- [15] LUO Y H, SONG H J, WANG R, et al. Arc FMCW SAR and applications in ground monitoring[J]. IEEE Transactions on Geoscience and Remote Sensing, 2014, 52(9): 5989-5998.

联合 ALE 与 TA-HA-Resnet 的舰船辐射噪声识别方法

陈祥毅^{1,2,3}, 张海刚^{1,2,3}, 钟玮琪^{1,2,3}, 李 坤^{1,2,3}

1. 哈尔滨工程大学 水声技术全国重点实验室, 哈尔滨 150001

2. 海洋信息获取与安全工信部重点实验室(哈尔滨工程大学), 哈尔滨 150001

3. 哈尔滨工程大学 水声工程学院, 哈尔滨 150001

摘要: 舰船辐射噪声中蕴藏着丰富的线谱信息, 这些特征对于舰船辐射噪声识别至关重要。传统基于深度学习的舰船辐射噪声识别方法, 主要关注于优化网络算法来提升识别精度, 却忽略了线谱及其谐波特征分量的利用。为了更有效地提取和利用线谱特征, 提出了一种名为 ALE-TA-HA-Resnet 的框架, 该框架通过增强线谱特征来提高舰船辐射噪声识别精度。该框架采用自适应线谱增强 (ALE) 技术对原始信号进行处理, 突出线谱特征, 并将其转化为时频图像, 方便网络捕获线谱特征的细节信息。引入两种线谱注意力机制: 基于时间注意力的 TA (temporal attention) 和基于谐波注意力的 HA (harmonic attention)。TA 可以聚焦于时频图像中的关键线谱特征, 而 HA 则能够识别并增强频带间谐波关系。通过将 TA 和 HA 并联融入 Resnet 网络中, 能够更有效地提取线谱特征, 并提升识别精度。在 ShipsEar 数据集上的实验结果表明, ALE-TA-HA-Resnet 模型在 5 类数据的识别中取得了 98.34% 的准确率, 比普通的 Resnet 模型提升了 6.73%, 证明了该框架在提高舰船辐射噪声识别精度方面的有效性。

关键词: 舰船辐射噪声识别; 深度学习; 自适应线谱增强; 基于 TA 和 HA 的线谱注意力机制

文献标志码: A

中图分类号: TP391.42

Ship Radiated Noise Recognition Method Based on ALE and TA-HA-Resnet

CHEN Xiangyi^{1,2,3}, ZHANG Haigang^{1,2,3}, ZHONG Weiqi^{1,2,3}, LI ZHEN^{1,2,3}

1. National Key Laboratory of Underwater Acoustic Technology, Harbin Engineering University, Harbin 150001, China

2. Key Laboratory of Marine Information Acquisition and Security (Harbin Engineering University), Ministry of Industry and Information Technology, Harbin 150001, China

3. College of Underwater Acoustic Engineering, Harbin Engineering University, Harbin 150001, China

Abstract: Ship radiated noise contains rich line spectral information, which is crucial for accurate ship identification. Traditional deep learning-based ship radiated noise recognition methods primarily focus on optimizing network algorithms to improve accuracy, neglecting the utilization of line spectral and harmonic components. To effectively extract and leverage these line spectral features, this paper proposes a framework named ALE-TA-HA-Resnet, which aims to enhance ship radiated noise recognition accuracy by enhancing line spectral features. Specially, this framework employs adaptive line spectral enhancement (ALE) to process the original signal, highlighting line spectral features and transforming them into time-frequency images, facilitating the network's capture of detailed line spectral information. Furthermore, two attention mechanisms based on line spectral are introduced: temporal attention (TA) and harmonic attention (HA). TA focuses on key line spectral features within the time-frequency images, while HA identifies and amplifies inter-band harmonic relationships. By integrating TA and HA into a parallel structure within the Resnet network, this framework can more effec-

作者简介: 陈祥毅 (2001—), 男, 硕士研究生, 研究方向为水声信号识别、海洋声场分析及应用; 张海刚 (1982—), 男, 博士, 教授, 研究方向为甚低频深海声传播、海洋声场分析及应用; 钟玮琪 (2000—), 女, 硕士研究生, 研究方向为水声信号识别; 李坤 (1996—), 男, 博士研究生, 研究方向为水声信号处理。

tively extract line spectral features and improve recognition accuracy. Experimental results on the ShipsEar dataset demonstrate that the ALE-TA-HA-Resnet framework achieves an accuracy of 98.34% in identifying 5 classes, surpassing the accuracy of the ordinary Resnet model by 6.73%. This proves the effectiveness of the framework in improving ship radiated noise recognition accuracy.

Key words: ship radiated noise recognition; deep learning; adaptive line spectrum enhancement; line spectral attention mechanism based on TA and HA

舰船辐射噪声是水声信号的主要来源之一，其信噪比受船只航行状态及环境噪声影响严重，增加了舰船辐射噪声的识别难度。因此在对此类数据的识别任务中，研究者通常会对数据进行预处理提取其特征，然后使用分类器进行数据的分类。在实际海洋环境中，传统特征提取方法如梅尔倒谱系数^[1]、高阶谱估计^[2]、希尔伯特黄变换^[3]、错误叠加算法等^[4]由于受到传播损失、海洋环境噪声等因素影响，难以保证辐射噪声数据信噪比^[5]。相比之下，深度学习手段能够学习特征信号的隐性数据特征，挖掘数据间的内在联系，建立数据至标签间稳健的映射关系，且不需要依赖过多的先验知识就能达到较为精确的拟合目标数据效果，具有良好的非线性数据拟合能力。在舰船辐射噪声识别任务中神经网络常用的特征提取方法是谱特征提取，如时频谱、梅尔倒谱、LOFAR 谱等。Baqar 等人^[6]使用了 LOFAR 谱作为 CNN 神经网络的输入，对 4 种不同的舰船辐射噪声进行了分类，达到了不错的识别准确率。于学洋等人^[7]使用 VGGish 神经网络获取基于梅尔滤波器组的特征矩阵。但仅依靠神经网络进行信号特征提取对识别精度的提升有限，若要进一步提升识别精度，可在数据预处理时增强利于神经网络学习的数据特征，如采用自适应线增强、自适应随机共振、隐马尔可夫模型、粒子滤波预检测跟踪等方法提高原始声信号的信噪比。或根据数据特点对网络结构进行优化。Li 等人^[8]中使用一种基于听觉启发的 CNN 神经网络对舰船辐射噪声进行识别，文中对辐射噪声线谱做一维卷积分解线谱，再对分解的线谱特征逐点卷积获取舰船辐射噪声的频域特征。赵乾坤等人^[9]中使用融合了注意力机制的 DNN 神经网络对数据特征进行二次提取。水声信号具有短时平稳的特点，该文基于该特点提取信号的帧级特征，并通过均值和标准差连接帧级特征转化为长时特征，有效提高了声信号的识别准确率。

水声信号信噪比较低，其连续谱易受到破坏，干扰模型的训练。窄带特征在频域中比宽带分量具有更高的功率且相对稳定，在低信噪比条件下通过特征增强技术可以恢复窄带线谱特征，从而提升网络的识别性能。为提升识别精度，本文从增强线谱特征的角度改进了舰船辐射噪声识别方法，采用自适应线谱增强（adaptive line enhancement，ALE）技术来增强线谱特征并抑制加性宽带噪声，从而提高线谱的信噪比^[10]。同时在网络中添加两种线谱增强注意力机制：时间注意机制（temporal attention，TA）和諧波注意力机制（harmonic attention，HA）^[11]捕捉线谱在时间维度上的平均特征及频带内諧波关系，使得网络能够更好地关注到线谱特征及频带间諧波关系。实验结果表明，经过 ALE 及两种线谱注意力机制处理的网络模型对舰船辐射噪声的判别准确率显著提高。

1 自适应线谱增强

自适应滤波由 Wiener 滤波和 Kalman 滤波等传统线性滤波方法发展而来，无需依赖先验知识即可在未知条件下实现最佳滤波^[12]。ALE 算法利用线谱分量与加性宽带白噪声在时间相关性上的差异进行线谱增强，能够有效抑制宽带噪声分量，并自适应恢复和增强被宽带噪声掩盖的线谱特征。舰船辐射噪声的线谱特征是船只判别的主要依据，ALE 能够有效增强其窄带线谱特征，提升网络训练样本质量^[13]。

假设被动声呐接收信号包括线谱分量 $s(k)$ 和高斯白噪声 $n(k)$ ，则 ALE 的输入信号离散形式可以表示为：

$$x(k) = s(k) + n(k) = \sum_{i=1}^M A_i \sin(2\pi f_i k + \varphi_i) + n(k) \quad (1)$$

其中， A 代表线谱分量的幅值， f 代表线谱分量的频率， φ 代表线谱分量的初始相位， M 表征线

谱个数。经过自适应线谱增强器的处理，输出信号可以表示为：

$$y(k) = \mathbf{w}^T(k)x(k-\Delta) \quad (2)$$

其中， $\mathbf{w}(k)$ 是 ALE 的自适应滤波器权值向量， $x(k-\Delta)$ 为滤波器时延 Δ 个单位的输入序列。

2 识别算法

本文模型依靠卷积神经网络 CNN 的变种模型 Resnet^[14]搭建，用于提取信号特征并作为分类器。同时为进一步提升线谱特征，增加网络识别正确率，在网络中加入两种针对线谱的注意力机制，使之能更好地学习线谱特征，实验总体框架如图 1 所示。

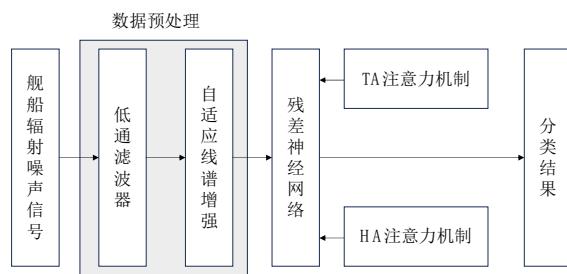


图 1 基于 ALE 和 TA-HA-Resnet 框架

2.1 注意力机制

舰船辐射噪声调制谱主要由离散频点的线性窄带线谱和连续谱叠加而成。同种船只在运行时线谱在频率上的波动较小，是辨别船只类型的主要依据^[15]。因此针对时频图像中的线谱特征，本文将 TA 和 HA^[11]两种线谱注意力机制融入 Resnet 中。输入注意力机制层的数据可以表示为 $A_{C,H,W}$ ， C 、 H 、 W 表示数据通道、高度及宽度，高度和宽度分别为频率和时间两个维度。

2.1.1 时间注意力机制 TA

线谱特征在频域范围内通常较为稀疏，在时频空间内直接对数据进行卷积计算容易引入过量噪声，TA 注意力机制旨在强化线谱特征，评估频带中谱特征在时间维度上的平均属性与数据像素间的相关性。其运算过程可以表示为：

$$TA = \text{Sigmoid}\left(\frac{1}{C} \sum_{C=1}^C \left(\left(\frac{1}{N} \sum_{w=1}^N Q_{C,H,W} \right) K_{C,H,W} \right) \cdot V_{C,H,W}\right) \quad (3)$$

其中， $Q_{C,H,W}$ 、 $K_{C,H,W}$ 、 $V_{C,H,W}$ 表示输入图像矩

阵 $A_{C,H,W}$ 经过卷积核尺度为 1×1 的卷积运算后的特征矩阵。为提取线谱在时间尺度上的平均特征， $Q_{C,H,W}$ 矩阵在时间维度上取平均并与 $K_{C,H,W}$ 矩阵进行 Hadamard 积以加强原有线谱特征。随后将 $A_{C,H,W}$ 与经过线谱注意力机制运算的矩阵相加以保证原始谱特性的保留。

2.1.2 谐波注意力机制 HA

舰船辐射噪声线谱由基波及高阶谐波构成，同种声源的谐波间存在明显的关联性，其谐波在时间维度上通常同时出现或同时消失，利用这种谐波关系可以帮助网络更好地学习船只特性。卷积运算中这种关系难以被捕捉，而 HA 注意力机制能够很好地捕捉谐波带间的关联性。其计算过程可以表示为：

$$HA = V_{C,H,W} \times \text{Sigmoid}(QT_{H,CW} \times KT_{CW,H}) \quad (4)$$

其中， $QT_{H,CW}$ 、 $KT_{CW,H}$ 为经过维度变换后的 $Q_{C,H,W}$ 、 $K_{C,H,W}$ 矩阵，在 HA 注意力机制中对形状重塑后的 $Q_{C,H,W}$ 、 $K_{C,H,W}$ 矩阵进行 Kronecker 积可以将线谱间的强能量关系及频带间线谱关系的特征映射到新的数据单元上，然后将经 HA 注意力机制模块处理的特征矩阵与原始图像矩阵相加以得到包含线谱相互关系的特征矩阵。

2.1.3 注意力机制连接

HA 注意力机制具有优异的频带内谐波关系捕捉能力，但运算过程中在时间维度上进行矩阵重塑忽略了数据的时间变化关系，易导致时间维度特征的丢失。TA 注意力机制能够捕捉线谱在时间维度上的平均特征，拓展时间维度感受野。因此将两种注意力机制并联连接能够在拓展感受野的同时使得网络能够更好地关注到线谱特征及频带间谐波关系，TA-HA 注意力机制运算流程如图 2 所示。

3 实验与结果分析

3.1 数据预处理

本文选用 ShipsEar 数据集作为实验数据来源。该数据集采集于 2012 年秋季到 2013 年夏季西班牙西北部大西洋沿岸 Vigo 港口附近海域。数据的观测时长在 15 s 至 10 min 不等，数据集根据噪声类型分为 A、B、C、D、E 共 5 类，其中包含 4 类不同大小的船只及背景噪声。船舶类型列于表 1，数据保留 52 734 Hz 的原始采样率。

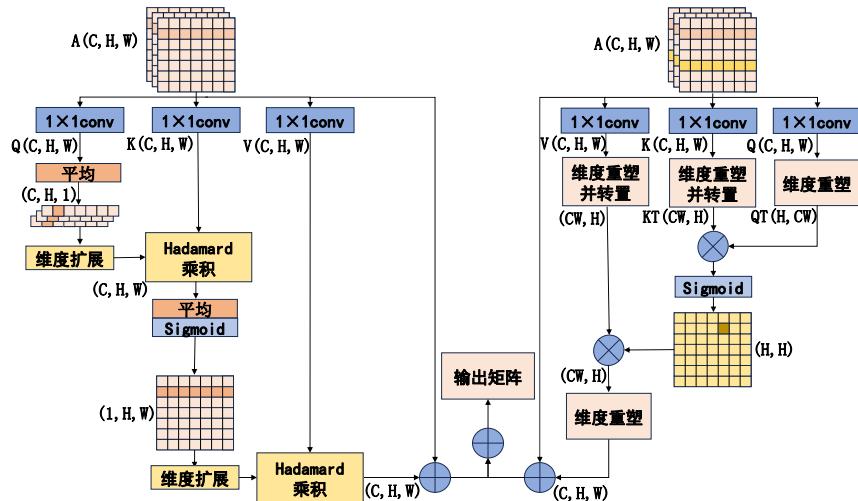


图 2 TA-HA 注意力机制运算流程

表 1 五类舰船类型

类别	舰船类型	训练样本数 /个	测试样本数 /个
A	渔船、拖网渔船、贻贝 船、拖船和挖泥船	556	139
B	摩托艇、领航船和帆船	447	112
C	客轮	1 286	322
D	远洋客轮和滚装船	755	189
E	背景噪音记录	334	84

舰船辐射噪声特征集中于低频段,因此首先对收集到的水声信号进行低通滤波滤除 4 kHz 以上的高频分量。其后对经过滤波处理的水声信号使用 ALE 增强线谱特征。其中滤波器阶数设置为 1 024 阶,延时设置 128 个采样点。图 3、图 4、图 5 分别为原始水声信号时域波形图、水声信号的时频谱图、水声信号经线谱增强后的时频谱图。由图 4、图 5 可见,经线谱增强处理后水声信号中的线谱分量得到了良好的保留,且噪声分量得到了有效的抑制,证明了 ALE 算法的有效性。

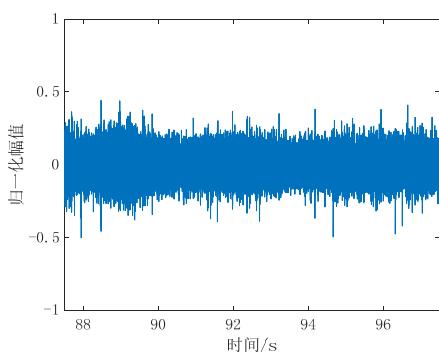


图 3 原始水声信号时域波形

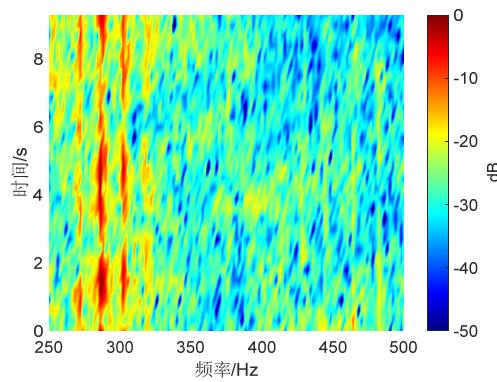


图 4 原始水声信号时频谱

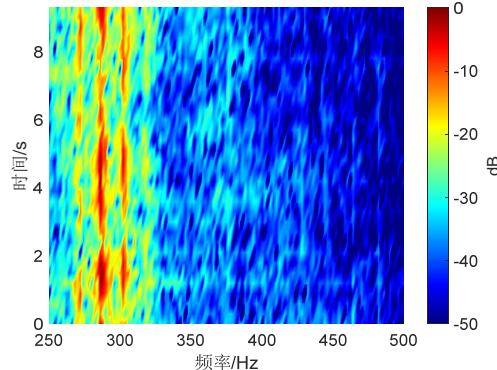


图 5 水声信号经自适应线谱增强后的时频谱图

为消除振幅差异影响、解决原始数据样本数量不足、信号的截止时间和最大振幅不一致的问题,对数据进行归一化和时间切片,将信号调整为具有相同时间尺度和幅度尺度的标准化样本。每个样本持续时间为 10 s,样本间重叠率为 75%。网络训练前还需对预处理后图像做标准化处理,标准化函数采用 Torchvision 库中的 transforms.Normalize 函数,具体参数为 [0.485, 0.456, 0.406], [0.229, 0.224,

0.225]。为减轻网络计算量，图片尺寸裁剪为 128×128 大小。最后按照4:1比例随机取样，将数据分为训练集和测试集，以降低训练集和测试集样本的相关性。

3.2 实验设置

本文使用Resnet18搭建网络模型，在snet.BasicBlock1后插入并联的两种注意力机制模块。为防止模型过拟合，在网络线性层前增加Dropout层(Dropout概率0.4)，损失函数为交叉熵损失函数，学习率设为 $1E-4$ ，批训练次数为32，使用Adam作为网络优化器进行梯度优化，训练epochs为20次。为评估本文提出框架的有效性，设置Resnet18、TA-HA-Resnet18、ALE-Resnet18 3组对照组进行消融实验，在测试集对比其损失值及分类准确率。

3.3 实验结果

实验结果表明，本文提出的ALE-TA-HA-Resnet框架在训练开始时损失值曲线快速收敛，随后趋于稳定，其损失曲线如图6所示。经20轮训练后，该框架在测试集上的准确率达98.34%，平均损失为0.061 46，与传统的Resnet框架相比，本文提出的框架预测准确率提升了6.73%。消融实验结果表明，相比于Resnet，ALE-Resnet及ALE-TA-HA-Resnet分类准确率更高，其中ALE-TA-HA-Resnet的分类准确率优于ALE-Resnet。证明ALE

表2 不同模型识别性能对比

方法	准确率/%	平均损失
Resnet	91.61	0.447 94
ALE-Resnet	94.97	0.216 06
TA-HA-Resnet	95.51	0.176 06
ALE-TA-HA-Resnet	98.34	0.061 46

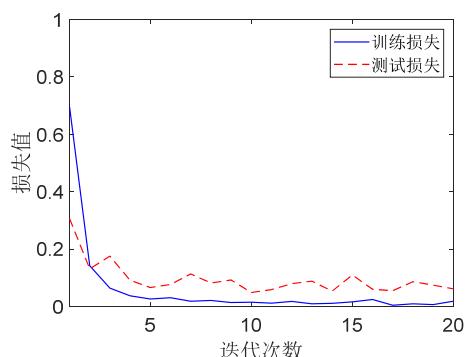


图6 损失曲线

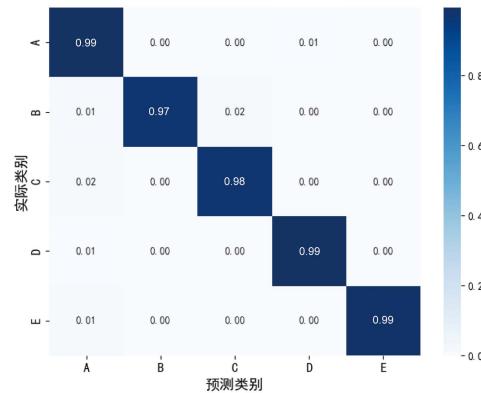


图7 混淆矩阵

预处理及TA-HA注意力机制能使网络更好地学习线谱特征。本文提出的ALE-TA-HA-Resnet框架能够很好地学习舰船辐射噪声中的线谱特征，具有较高的识别精度。

4 结束语

本文从增强线谱特征的角度提出了一种ALE-TA-HA-Resnet框架用于识别舰船辐射噪声。该框架对低通滤波后的音频数据进行ALE预处理，使得时频图像中线谱特征显著且背景噪声进一步降低。此外，在Resnet网络中引入TA和HA两种注意力机制，进一步在端对端学习中增强了有助于识别精度提升的线谱特征。通过实验的对比验证，本文提出的框架准确率相比Resnet提高了6.73%，有效提升了舰船辐射噪声识别精度，为此类识别任务提供了新思路。然而，目前舰船辐射噪声在低信噪比情况下的识别仍存在一定的困难，不能很好地保证识别精度，未来对舰船辐射噪声的研究动向可以围绕此问题展开。

参考文献：

- [1] 刘聪, 韩东, 张欣洋, 等. 面向原始辐射噪声的水声目标识别研究[J]. 电声技术, 2023, 47(8): 30-37.
- [2] 吴国清, 李靖, 陈耀明, 等. 舰船噪声识别(I): 总体框架、线谱分析和提取[J]. 声学学报, 1998, 23(5): 394-400.
- [3] YANG S, XUE L, HONG X, et al. A lightweight network model based on an attention mechanism for ship-radiated noise classification[J]. Journal of Marine Science and Engineering, Multi-disciplinary Digital Publishing Institute, 2023, 11(2): 432.

(下转第179页)

自监督学习多普勒谱重构的海杂波抑制方法

赵振方¹, 王文光¹, 李一鸣², 吴巍³

1. 北京航空航天大学 电子信息工程学院, 北京 100191

2. 中国信息通信研究院 泰尔系统实验室, 北京 100083

3. 中国人民解放军海军工程大学 兵器工程学院, 武汉 430033

摘要: 海上目标的稳定检测一直是雷达技术的研究热点。海杂波抑制可以提升雷达对海上弱目标的检测能力, 然而, 海杂波的强时变性与空变性使得传统的海杂波抑制方法存在模型与实际观测数据失配的问题, 模型的先验假设难以满足。为此, 提出了一种数据驱动的海杂波抑制方法, 所提方法无需目标回波样本作为标签, 仅使用海杂波训练自编码器重构距离-多普勒谱, 与实测回波对消来抑制海杂波, 从而检测海上目标。基于实测数据对所提方法进行了实验验证, 实验结果表明, 所提出方法可以显著抑制高强度、非均匀的海杂波背景, 在稳定检测目标的同时可以有效减少虚警, 是一种有效的海杂波抑制方法。

关键词: 海杂波抑制; 多普勒谱; 自监督学习; 自编码器; 目标检测

文献标志码: A

中图分类号: TP391

Sea Clutter Suppression via Self-Supervised Learning Doppler Spectrum Reconstruction

ZHAO Zhenfang¹, WANG Wenguang¹, LI Yiming², WU Wei³

1. School of Electronic and Information Engineering, Beihang University, Beijing 100191, China

2. CTTL-Systems, China Academy of Information and Communications Technology, Beijing 100083, China

3. College of Weaponry Engineering, Naval University of Engineering, Wuhan 430033, China

Abstract: The stable detection of maritime targets has always been a hotspot of radar technology research. Sea clutter suppression can improve the detection ability of radar on dim marine targets. However, due to the strong temporal and spatial variability of sea clutter, the traditional sea clutter suppression method has the problem of mismatch between the model and the actual observation data, and the prior assumptions of the model are difficult to meet. Therefore, a data-driven sea clutter suppression method is proposed in this paper. The proposed method does not need the target echo sample as a label, and only uses the sea clutter to train auto-encoder to reconstruct the range-Doppler spectrum and cancel with the measured echo to suppress the sea clutter, so as to detect the marine targets. Based on the measured data, the proposed method is verified by experiments. The experimental results show that the proposed method can significantly suppress the high-intensity and non-uniform sea clutter background, and effectively reduce false alarms while stably detecting the target. The proposed method is an effective sea clutter suppression method.

Key words: sea clutter suppression; Doppler spectrum; self-supervised learning; auto-encoder; target detection

在雷达对海目标探测时, 海杂波是普遍存在的环境杂波, 其由海面的强后向散射形成, 海杂波的

存在显著降低了目标检测的概率。对于雷达截面积 (radar cross section, RCS) 较小的海上弱目标,

基金项目: 国家自然科学基金 (62073334)。

作者简介: 赵振方 (2000—), 男, 硕士研究生, 研究方向为海上目标检测技术; 王文光 (1975—), 男, 教授, 研究方向为目标检测、定位等; 李一鸣 (1977—), 男, 硕士, 研究方向为目标检测、定位等; 吴巍 (1981—), 男, 副教授, 研究方向为海上目标检测、跟踪等。

在较高海况时，其回波能量远低于海杂波，难以进行有效检测。因此，有效的海杂波抑制对于海上目标探测十分重要。

传统的海杂波抑制方法从海杂波与目标的幅度信息、谱信息出发，尝试在不同域上将目标和海杂波进行区分，并基于先验假设，设计有效的海杂波抑制算法。其中具有代表性的一些方法包括奇异值分解（singular value decomposition, SVD）^[1]、子空间分解^[2]、自适应滤波^[3]、AMTI（adaptive moving target indication）、STAP（space time adaptive processing）^[4]等。研究表明，在高频雷达下，海杂波的能量主要集中于 Bragg 散射分量处，基于 SVD 的海杂波抑制方法通过对距离单元内接收回波的 Hankel 矩阵进行分解，抑制海杂波的 Bragg 分量，从而实现海杂波抑制^[1]。将 SVD 与分数阶傅里叶变换（fractional Fourier transform, FrFT）结合，SVD-FrFT 方法也被提出用于抑制海杂波，其通过将回波数据变换到 FrFT 域，构造 Hankel 矩阵，再通过 SVD 实现海杂波抑制^[5]。基于子空间分解的海杂波抑制根据实测数据构造协方差矩阵，并进行特征值分解得到海杂波抑制子空间，将信号正交投影到海杂波抑制子空间，实现海杂波抑制。自适应滤波对待处理单元相邻距离单元的海杂波多普勒谱进行估计，设计相应的滤波器抑制海杂波^[4]。多通道的自适应滤波通过在海杂波的离散频率处构建多个具有尖锐陷波边缘的滤波器，从而实现杂波抑制并保留目标^[6]。MTI^[7]是在目标检测中被广泛应用的杂波抑制技术，其通过脉冲对消抑制零多普勒频率处的海杂波。然而，在风和海浪运动的影响下，海杂波的多普勒谱中心频率常常偏离零频，导致基于 MTI 的杂波抑制方法的性能下降。AMTI 通过估计海杂波多普勒谱的中心频率，采用自适应权值的方法实现了对多普勒动态范围较大的海杂波的抑制。STAP 综合考虑回波时域和空域的信息来抑制海杂波，为海杂波抑制提供了新的思路和方法^[8]。除此之外，一些基于海杂波的分形特征^[9-10]和时频特征^[11-12]的海杂波抑制方法也相继被提出。然而，以上方法大都依赖于对海杂波的先验假设和建模估计。在实际应用中，海杂波的非平稳性和强时变性使其特性偏离经验模型，也常与先验假设不符，因此对传统的海杂波抑制方法处理后

的回波进行目标检测时，依然存在大量残余海杂波造成的虚警。

随着机器学习技术的发展，基于数据驱动的海杂波抑制方法逐渐受到关注，深度神经网络开始被应用于海杂波抑制。从海杂波的时序特性和相位出发，复值 LSTM（long short-term memory）网络被用于在时域上预测海杂波，通过将实测数据与复值 LSTM 网络的预测值对消来抑制海杂波，取得了较好的抑制效果^[13]。除此之外，复值 U-Net 网络也被用于非均匀海杂波的特征提取和抑制^[14-15]。图像到图像的复值 U-Net 网络将距离-多普勒谱的复值图像直接作为输入，并将一致性损失函数用于距离-多普勒域抑制海杂波，并最大程度保留目标信号，在具有较低计算复杂度的同时获得了较高的信杂比提升^[15]。基于生成对抗网络（generative adversarial networks, GAN）的海杂波抑制方法^[16]将杂波抑制问题转化为含杂波雷达数据到无杂波雷达数据的转换，并基于 CycleGAN 网络和一致性损失训练得到了海杂波抑制网络，取得了良好的杂波抑制性能。然而，目前基于深度神经网络的海杂波抑制方法大多依赖于大量杂波样本和相应的无杂波样本作为训练的输入与标签，且目标样本常常是仿真得到的合作目标。而在海面回波中，海杂波存在于观测时间内的所有距离单元上，而目标回波只存在于少数几个距离单元内，目标回波相比于海杂波具有稀疏性，且实测数据中的海上目标是非合作的。因此，从实测数据中获取大量的目标回波样本构建训练标签是困难的，且仿真得到的合作目标的特性也与实际的海上目标存在较大差异。而大量海杂波数据的训练样本更为容易获取。从这个角度出发，使用自监督学习的方法可以避免监督学习的样本类别不平衡问题，更适合于海杂波抑制和海上目标检测。相比于时域回波，距离-多普勒域的目标和海杂波特性差异更加明显，且经过多脉冲积累后信杂比的提升更有利后续的目标检测，因此在距离-多普勒域抑制海杂波并检测目标是一种合理且有效的方法。

本文将研究通过自监督学习提取海杂波的多普勒域特征，并重构海杂波多普勒谱的方法，为海杂波抑制和目标检测提供新的思路和方法。

1 海杂波距离-多普勒域数据集构建

对于工作在凝视模式下的雷达, 经脉冲压缩后接收到的 N 个相干脉冲的回波可表示为矩阵 $X \in \mathbb{C}^{N \times M}$, 其中 M 为接收回波的距离单元数。对于接收到的回波, 使用滑窗从原始回波中采集得到每帧样本:

$$X_k(n, m) = X(s \cdot (k-1) + 1 : s \cdot (k-1) + n, 1 : m) \quad (1)$$

其中, s 为样本间隔, n 为样本长度, m 为每帧样本的距离单元数。对于滑窗得到的每帧样本进行多普勒处理, 实现相参积累:

$$X_k(f_d, m) = \left| \sum_{i=1}^n X_k(i, m) \exp\left(-j \frac{2\pi}{n} f_d i\right) \right| \quad (2)$$

其中, f_d 为多普勒频率。为了提高网络模型的收敛速度和重构精度, 对相参积累后的得到的每帧样本进行最大值归一化:

$$X_{nor, k} = \frac{X_k}{\max(X_k)} \quad (3)$$

其中, $\max(\cdot)$ 表示求最大值。经归一化后, 每帧样本的值分布在 0 到 1 之间, 更有利于网络的训练。

2 基于多普勒谱重构的海杂波抑制方法

深度自编码器由一个编码器和一个解码器组成, 使用预处理后得到的样本 $X_{nor, k}$ 训练的编码器习得了海杂波多普勒域的特征表示, 能够将输入数据转化为特征空间中的压缩编码 $\chi = E_\eta(X_{nor, k})$, 解码器根据编码得到的特征尽可能准确地重构输入, 自编码器的输出 $\hat{X}_{nor, k} = D_\delta(E_\eta(X_{nor, k}))$ 。其中 η 和 δ 分别为编码器 $E_\eta(\cdot)$ 和解码器 $D_\delta(\cdot)$ 的可学习参数。为了使得输出与输入的误差最小, 训练的损失函数使用基于二范数的均方误差 L_{MSE} , 其形式为:

$$L_{MSE} = \frac{1}{K} \sum_{k=1}^K (X_k - \hat{X}_k)^2 \quad (4)$$

当输入为海杂波数据时, 训练后的自编码器可以以较小的误差重构海杂波多普勒谱, 并通过对消来抑制海杂波。当输入数据中包含目标回波时, 由于目标和海杂波在距离-多普勒域的特征差异和可分性, 海杂波被准确重构, 而目标处则产生大的重

构误差, 因此经过对消后海杂波被有效抑制, 而目标信号得到保留, 因此基于多普勒谱重构的海杂波抑制方法可以显著提升目标检测的信杂比。

图 1 所示为基于多普勒谱重构的海杂波抑制与目标检测方法的实现流程。海杂波抑制后的回波信号的信杂比得到显著提升, 因此可以在较低的虚警率下实现海上目标的稳定检测。对于海杂波抑制后的回波, 在距离-多普勒域使用二维 CFAR 算法实现目标检测。



图 1 基于多普勒谱重构的海杂波抑制方法

图 2 所示为所提出方法的自编码器的网络结构, 包含了一个编码器和一个解码器。

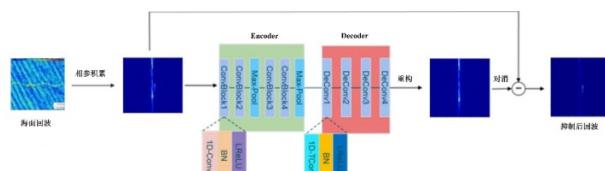


图 2 基于自编码器结构的海杂波抑制

3 实验结果与分析

为验证所提出方法对于海杂波的抑制效果, 本章基于实测数据, 从自编码器的结构和参数设计、海杂波抑制性能、目标检测性能等方面, 对所提方法进行验证。

实验所用数据为 Fynmeet 雷达于南非西南海岸线采集的编号为 TFC17-006 的 CSIR 实测数据, 回波包含了海杂波以及运动的小船, 实验数据的具体参数如表 1 所示。

表 1 TFC17-006 参数说明

数据参数	值	数据参数	值
目标类型	小船	持续时间/s	111.020 4
载频/GHz	9	擦地角	0.501~0.56 deg.
PRF/kHz	5	平均风速	15.2 kts, 249.7 deg. N.
距离分辨率/m	15	阵风风速	26 kts, 264 deg. N.
起始距离/m	6 600	海浪高度	2.35 m, 144.9 deg. N.

实验所用数据的距离-时间图如图 3 所示。

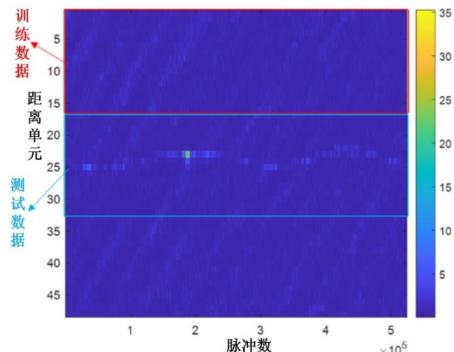


图 3 TFC17-006 数据的距离-时间图

在观测时间内，目标位于第 23 至第 25 距离单元内。图 3 中红色框标出的数据是训练数据，为 1~16 距离门内，持续时间为 105 s 海杂波。蓝色框标出的数据为含运动目标的测试数据，范围为 17~32 距离门内的回波。使用第 1 章所述方法，对训练数据进行采样得到海杂波距离-多普勒域样本，并将样本划分为训练集和验证集。对测试数据进行采样得到测试集。

3.1 自编码器网络结构

对于输入的距离-多普勒域回波，使用多通道一维卷积提取不同距离单元的多普勒谱特征，得到编码结果，并根据编码得到的特征重构原始输入。由于自编码器的超参数会对实验结果产生影响，因此需要详细给出实验时的超参数设置。所提出的海杂波抑制方法是基于自编码器结构的，包括编码器和解码器。表 2、表 3 分别给出了编码器和解码器每层的超参数设置。

表 2 编码器超参数设置

层数	Kernel Size	Stride Size	Channel
Conv1	32	2	128
Conv2	32	2	256
Conv3	16	2	512
Conv4	16	2	512

表 3 解码器超参数设置

层数	Kernel Size	Stride Size	Channel
Deconv1	4	4	256
Deconv2	2	2	128
Deconv3	4	4	128
Deconv4	2	2	16

3.2 海杂波抑制性能验证

本节将使用包含运动目标的实测回波，验证所

提出方法对于回波信杂比的改善效果。数据预处理阶段，设置采样间隔 $s = 64$ ，每帧样本的距离单元数 $m = 16$ ，脉冲数 $n = 512$ ，因此，自编码器的输入为通道数 $C = 16$ ， $L = 512$ 的多普勒谱。在训练阶段，使用 Adam 优化器，设置学习率，batch size 大小为 64，并训练 1 000 个 epoch，得到用于海杂波抑制的自编码器模型。

图 4 为使用所提出方法对于测试数据进行海杂波抑制的实验结果。其中图 4 第一行所示为运动速度较慢，目标多普勒谱落入海杂波多普勒区域内情况下的目标 1 海杂波抑制效果。第二行为运动速度较快，目标多普勒谱远离海杂波区域情况下的目标 2 的海杂波抑制结果。图中第一列为包含目标的海杂波多普勒谱，第二列为重构多普勒谱，第三列为抑制后的结果，图中红框为目标所在位置。

图 4 所示结果表明，自编码器通过自监督学习有效学习了海杂波的 RD 域特征，故可以准确回归并重构海杂波的多普勒谱。即使在含有目标的数据中，也有效实现了海杂波重构，从而实现海杂波抑制。所提出的海杂波抑制方法可以有效抑制高强度、非均匀的海杂波。对于不同运动状态目标的实验结果表明，所提出方法可以在显著抑制海杂波的同时保留海上目标，不仅可以在目标与海杂波 RD 域可分的情况下抑制海杂波，还可以在慢速小目标情况下有效抑制海杂波，具有较强的鲁棒性。

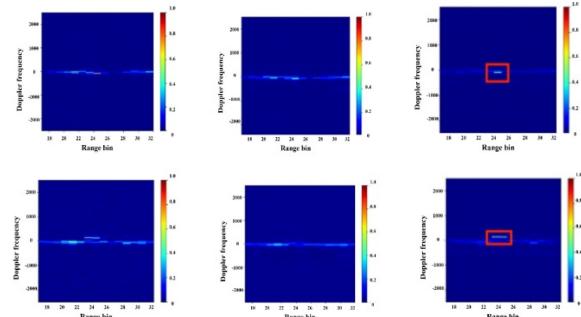
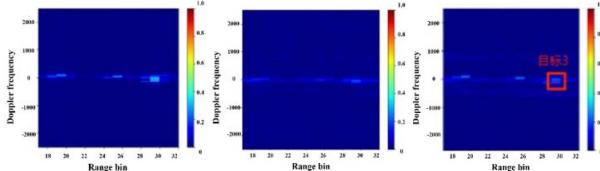


图 4 基于多普勒谱重构的海杂波抑制实验结果

为了验证所提出方法的泛化能力，实验还使用了与 TFC17-006 数据的采集时间和环境参数不一致的，编号为 TFC17-002 的数据。图 5 所示为使用 TFC17-006 的海杂波训练后的模型对于 TFC17-002 数据的抑制效果，将目标编号为 3，从左到右分别为 RD 谱、重构 RD 谱和抑制后结果。可以看出，所提出方法对于观测条件不同的数据也



(a) 目标 3 RD 谱 (b) 目标 3 重构结果 (c) 抑制结果

图 5 TFC17-002 数据的海杂波抑制实验结果

具有较好的海杂波抑制能力。

表 4 给出了所提出方法对于不同测试数据的信杂比提升情况。由表 4 所示结果可知, 基于多普勒谱重构的海杂波抑制方法可以显著抑制海杂波背景, 提升观测数据的信杂比, 且对于不同观测条件下的数据具有一定泛化能力, 是一种有效的海杂波抑制方法。

表 4 海杂波抑制前后信杂比对比 单位: dB

实验数据	TFC17-006	TFC17-002
抑制前平均	17.25	14.59
抑制后平均	29.70	24.53
SCR 提升	12.45	9.94

3.3 海杂波抑制后目标检测性能验证

海杂波抑制后的数据提升了信杂比, 这里以目标检测为应用, 并与基于小波分解的目标检测方法进行对比, 验证所提出方法可以在稳定检测目标的同时, 显著减少虚警数目。

对于海杂波抑制后的数据, 使用双参数的 CFAR 检测进行目标检测实验, 设置背景窗和保护窗的大小后, 基于海杂波高斯分布假设, 计算背景窗样本的标准差和均值, 检测的阈值可以由下式计算得到:

$$th_d = \text{mean}(c) + cof \cdot \text{std}(c) \quad (5)$$

其中 $\text{std}(\bullet)$ 为取标准差运算, $\text{mean}(\bullet)$ 为取均值运算。 c 为背景点的集合, cof 为检测系数, 以此来对检测阈值进行调节。

得到的检测结果如图 6 所示, 图中红框为真实目标位置。其中图 6 第一行使用所提出方法进行海杂波抑制后对目标 1、目标 2、目标 3 的检测结果, 图 6 第二行给出了基于小波分解方法抑制海杂波后的目标 1、目标 2、目标 3 检测结果。由图 6 所示结果可知, 在相同的积累点数下, 因海杂波多

普勒动态范围较大, 导致基于小波分解的检测方法提取的目标中, 存在大量由海杂波所造成的虚警。而基于多普勒谱重构的方法通过数据驱动的方式学习到了海杂波的多普勒域特征, 准确地回归了海杂波的多普勒谱, 有效抑制了海杂波, 有效减少了海杂波造成的虚警。

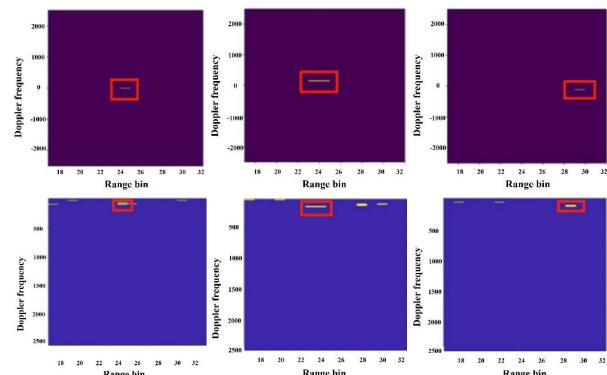


图 6 海杂波抑制后的目标检测结果

4 结束语

针对传统的基于先验假设和模型的海杂波抑制方法与实测数据失配的问题, 提出了一种基于数据驱动的自监督学习海杂波抑制方法。所提出方法只需使用杂波数据进行训练, 学习海杂波多普勒谱特征并重构输入, 通过将重构结果与输入对消来抑制海杂波, 提升信杂比, 进而提升海上目标检测性能。该方法无需构造无杂波的目标样本作为标签, 避免了监督学习应用于海上目标探测存在的类别不平衡问题。实验结果验证了所提出方法可以显著提升回波信杂比, 并可以稳定检测海上目标的同时大量减少虚警数。后续的工作主要将聚焦于提升方法对于不同观测条件下数据的泛化能力以及网络超参数的调优, 尝试取得更高的信杂比改善以及更强的泛化能力。

参考文献:

- [1] POON M W Y, KHAN R H, LE-NGOC S. A singular value decomposition (SVD) based method for suppressing ocean clutter in high frequency radar[J]. IEEE Transactions on Signal Processing, 1993, 41(3): 1421-1425.
- [2] WANG Z, WANG Y, DUAN K, et al. Subspace-augmented clutter suppression technique for STAP radar[J].

- IEEE Geoscience and Remote Sensing Letters, 2016, 13(3): 462-466.
- [3] SHI Y L, SHUI P L. Target detection in high-resolution sea clutter via block-adaptive clutter suppression[J]. IET Radar, Sonar & Navigation, 2011, 5(1): 48-57.
- [4] LEI Z, JINPING S, XIAOJIAN X, et al. Space-time adaptive processing for sea clutter and jamming suppression in radar seekers[C]//2010 2nd International Conference on Signal Processing Systems. IEEE, 2010, 3: V3-600-V3-604.
- [5] CHEN Z, HE C, ZHAO C, et al. Using SVD-FRFT filtering to suppress first-order sea clutter in HFSWR[J]. IEEE Geoscience and Remote Sensing Letters, 2017, 14(7): 1076-1080.
- [6] LIU S, MA Y, HUANG Y. Sea clutter cancellation for passive radar sensor exploiting multi-channel adaptive filters[J]. IEEE Sensors Journal, 2018, 19(3): 982-995.
- [7] CAPON J. Optimum weighting functions for the detection of sampled signals in noise[J]. IEEE Transactions on Information Theory, 1964, 10(2): 152-159.
- [8] 褚天琦. 基于空时自适应处理的展宽海杂波抑制[D]. 哈尔滨: 哈尔滨工业大学, 2014.
- [9] DU G. Detection of sea-surface radar targets based on fractal model[J]. Electronics Letters, 2004, 40(14): 1.
- [10] GUAN J, LIU N, ZHANG J, et al. Multifractal correlation characteristic for radar detecting low-observable target in sea clutter[J]. Signal Processing, 2010, 90(2): 523-535.
- [11] SHI S N, SHUI P L. Sea-surface floating small target detection by one-class classifier in time-frequency feature space[J]. IEEE Transactions on Geoscience and Remote Sensing, 2018, 56(11): 6395-6411.
- [12] ZUO L, LI M, ZHANG X, et al. An efficient method for detecting slow-moving weak targets in sea clutter based on time-frequency iteration decomposition[J]. IEEE Transactions on Geoscience and Remote Sensing, 2012, 51(6): 3659-3672.
- [13] ZHU H, YU Z, YU J. Sea clutter suppression based on complex-valued neural networks optimized by PSD[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2022, 15: 9821-9828.
- [14] WANG Y, ZHAO W, WANG X, et al. Nonhomogeneous sea clutter suppression using complex-valued U-Net model[J]. IEEE Geoscience and Remote Sensing Letters, 2022, 19: 1-5.
- [15] QU Q, CHEN H, LIU W, et al. Data-driven sea clutter suppression via an image-to-image neural network and consistency loss[J]. IEEE Transactions on Aerospace and Electronic Systems, 2024.
- [16] PEI J, YANG Y, WU Z, et al. A sea clutter suppression method based on machine learning approach for marine surveillance radar[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2022, 15: 3120-3130.

(上接第 173 页)

- [4] 程星. 低信噪比水声信标信号处理方法研究及试验 [D]. 大连: 大连海事大学, 2018.
- [5] 李笑媛, 程晶, 那健. 一种基于 DSP 的声信标信号提取方法[J]. 舰船科学技术, 2014, 36(11): 125-127.
- [6] BAQAR M, ZAIDI S S H. Performance evaluation of linear and multi-linear subspace learning techniques for object classification based on underwater acoustics[C]// 2017 14th International Bhurban Conference on Applied Sciences and Technology (IBCAST). IEEE, 2017: 675-683.
- [7] 于学洋, 李淑秋, 宁江波, 等. 一种基于 VGGish 神经网络的水声目标识别方法[J]. 声学技术, 2024, 43(1): 30-37.
- [8] LI J, YANG H. The underwater acoustic target timbre perception and recognition based on the auditory inspired deep convolutional neural network[J]. Applied Acoustics, 2021, 182: 108210.
- [9] 赵乾坤, 刘峰, 梁秀兵, 等. 基于时延神经网络模型的舰船辐射噪声目标识别[J]. 应用声学, 2023, 42(5): 1033-1041.

- [10] 金盛龙, 迟骋, 李宇, 等. 稀疏驱动自适应线谱增强的水下目标谱熵检测[J]. 声学学报, 2021, 46(6): 1059-1069.
- [11] LIANG G, CHEN Y, WANG J, et al. Enhanced noise resilience in passive tone detection via broad-receptive field complex-valued convolutional neural networks[J]. The Journal of the Acoustical Society of America, 2024, 155(6): 3968-3982.
- [12] HAO Y, CHI C, QIU L, et al. Sparsity - based adaptive line enhancer for passive sonars[J]. IET Radar Sonar Navig., 2019, 13(10): 1796-1804.
- [13] 于学洋, 迟骋, 李淑秋, 等. 联合线谱增强与深度神经网络的水声目标识别[J]. 声学学报, 2024, 49(4): 656-663.
- [14] HONG F, LIU C, GUO L, et al. Underwater acoustic target recognition with resnet18 on shipsear dataset[C]//2021 IEEE 4th International Conference on Electronics Technology (ICET). IEEE, 2021: 1240-1244.
- [15] YANG X, MU L, LIU Y. Study on line-spectrum stability of ship underwater radiated noise[C]//China Ocean Acoustics (COA). IEEE, 2016: 1-5.

基于稀疏阵列矩阵补全的毫米波雷达二维 DOA 估计方法

汪济原¹, 杜 磊², 王彦平¹

1. 北方工业大学 信息学院, 北京 100144

2. 中国计量科学研究院 力学与声学计量科学研究所, 北京 100029

摘要: 在毫米波雷达的实际应用场景中, 由于阵元数量的限制, 常使用稀疏布阵的方法布置阵元, 相较于均匀阵列增大了阵列孔径和分辨能力。但是, 稀疏阵列对于接收信号的采样率不满足奈奎斯特采样定律, 这导致在进行波达方向估计时, 计算目标空间谱的旁瓣增高, 易受噪声影响并出现虚警现象。为了解决此问题, 针对方位向阵元均匀排布、俯仰向阵元稀疏排布的稀疏面阵, 提出了一种毫米波雷达多维目标参数处理方法, 并对原子范数最小化重构协方差矩阵波达方向估计算法进行改进, 实现精确提取目标的距离、方位角、俯仰角、速度等参数信息。该方法采用 2D-FFT 提取目标的距离、速度信息, 随后使用奇异值阈值方法进行粗略重构协方差矩阵, 再迭代多次原子范数最小化二次精细重构协方差矩阵实现虚拟阵元补全, 最后进行 2D-MUSIC 处理, 实现方位角和俯仰角的高精度波达方向估计。

关键词: 毫米波雷达; 稀疏阵列; 矩阵补全; DOA 估计; 原子范数; 目标参数处理

文献标志码: A

中图分类号: TN959.6; TN98

Millimeter-Wave Radar 2D DOA Estimation Method Based on Sparse Array Matrix Completion

WANG Jiyuan¹, DU Lei², WANG Yanping¹

1. School of Information Science and Technology, North China University of Technology, Beijing 100144, China

2. Division of Mechanics and Acoustics Metrology, National Institute of Metrology, Beijing 100029, China

Abstract: Due to the limitation on the number of array elements, sparse array configurations are widely used in practical millimeter-wave radar applications. Compared with uniform rectangular arrays, sparse arrays extend array aperture and resolution capability. However, sparse arrays do not satisfy the Nyquist sampling theorem for the received signal, leading to elevated sidelobes in the spatial spectrum when estimating the direction of arrival (DOA), making the system more susceptible to noise and leading to a higher likelihood of false alarms. This paper proposes a multidimensional target parameter processing method for millimeter-wave radar using a sparse planar array, where the azimuth elements are uniformly distributed and the elevation elements are sparsely distributed, and improves upon the atomic norm minimization reconstruction covariance matrix DOA estimation algorithm to accurately extract target parameters such as range, azimuth angle, elevation angle, and velocity. The approach first employs 2D-FFT to extract the target's range and velocity information, then uses singular value thresholding for the coarse reconstruction of the covariance matrix, followed by multiple iterations of atomic norm minimization-based fine reconstruction of the covariance matrix to achieve virtual array element completion. Finally, high-precision DOA estimation of azimuth and elevation angles is achieved using 2D-MUSIC.

Key words: millimeter-wave radar; sparse array; matrix completion; DOA estimation; atomic norm minimization; multi-dimensional target parameter processing

作者简介: 汪济原 (1999—), 男, 硕士研究生, 研究方向为雷达阵列信号处理; 杜磊 (1982—), 通信作者, 男, 副研究员, 研究生导师, 研究方向为运动学参数计量, E-mail: dulei@nim.ac.cn; 王彦平 (1976—), 通信作者, 男, 教授, 研究员, 研究方向为智能雷达、雷达形变监测、智能预警技术, E-mail: wangyp@ncut.edu.cn。

随着计算机科学和人工智能的发展，自动驾驶技术快速发展并成为当今汽车行业的热点^[1]。自动驾驶技术为交通运输带来了更高效、更安全、更环保的前景。毫米波雷达作为自动驾驶的重要传感器，广泛应用于汽车测速、倒车、目标检测等场景，其优势在于体积小、成本低、全天候工作、分辨率高等。对毫米波雷达捕获的雷达数据进行处理，获得车辆、行人等目标的多维运动参数，设计毫米波雷达多维目标参数处理算法成为研究热点^[2]。毫米波雷达目标参数主要包括目标的空间距离、方位、高度和速度，可以通过获取距离、速度、方位角以及俯仰角来计算。但是雷达原始数据较大、处理算法复杂；在复杂场景下，受杂波、噪声和低分辨率的影响，导致漏检和虚警问题，从而限制了其在复杂感知任务中的适用性。因此，为毫米波雷达设计相应的算法已成为研究热点。

近年来，毫米波雷达多维目标参数处理算法有了较大进展，研究人员在测距、测速、测角算法上设计、改进了多种算法。在距离参数提取方面，目前经典的测距算法主要有：基于傅里叶变换的 FFT 方法以及其改进算法如 ZFFT (Zoom-FFT) 等，还有学者将波达方向估计的超分辨算法应用于测距，例如 MUSIC 算法等。文献[3]对基于傅里叶变换的 chirp-z 测距算法以及基于压缩感知的超分辨测距算法进行改进，提升了算法的测量准确度并降低了算法处理复杂度。在文献[4]中，作者将改进 FFT 算法与 chirp-z 算法相结合，设计了 Fast/slow chirp 波形解决多普勒模糊的问题，实现了高精度的距离测量。

在速度参数提取方面，目前经典的测速算法主要有：差分相位测速法、互相关测速法、基于 keystone 变换测速法等。在文献[5]中，作者对传统距离向互相关测速方法进行改进，将距离单元的整数阶单元推广到了分数阶单元，提高了测速的精准度，但同时也大大增大了计算复杂度。在文献[6]中，作者在 keystone 变换的基础上，提出了 KT-VD-BS 算法，在低信噪比条件下，对信号相位进行恢复，解决了距离徙动和多普勒模糊问题，实现了高精度速度估计。

在角度参数提取方面，波达方向 (direction of arrival, DOA) 估计是处理毫米波雷达目标参数的

关键步骤之一^[7]，通过处理阵列接收信号获取目标的方位参数。目标波达方向估计的精准程度决定了估计目标空间位置的准确程度，是影响智能驾驶安全性和驾驶体验的关键因素。

在实际应用中，由于阵元数量的限制，在毫米波雷达阵列设计中常采用稀疏阵列，使用较少阵元进行高精度、高性能的 DOA 估计。适用于均匀线性阵列的 DOA 估计算法如子空间类 MUSIC (multiple signal classification) 算法等不能直接处理稀疏阵列接收信号，因此需要设计相应稀疏阵列 DOA 估计算法。近些年来，随着压缩感知和矩阵填充理论的发展，许多学者开发了许多基于低秩矩阵重构的稀疏阵列 DOA 估计算法^[8]。文献[9]针对嵌套阵列提出了一种基于 Toeplitz 重构的压缩感知 DOA 估计算法，该算法在对稀疏向量进行加权约束的条件下，将虚拟阵列信号重构为满秩协方差矩阵。然而，该算法需要在多快拍条件下才能获得较好 DOA 估计性能。文献[10]提出了一种改进的奇异值阈值 (singular value threshold, SVT) 算法，该算法引入惩罚机制并将软阈值运算应用于稀疏矩阵来处理核范数最小化的问题。然而，由于核范数是秩函数的松散逼近，得到的 SVT 解通常是次优的，这将影响后续的 DOA 估计结果。文献[11]通过构建选择矩阵来消除相位耦合的影响，并使用 l_1 -SVD 算法进行 DOA 估计，但该算法存在阵列孔径损失的问题。同时，文献[12]提出了一种基于协方差向量的稀疏感知算法，通过向量化协方差矩阵来获得等效均匀线性阵列接收信号的单快拍，该方法的计算量大、数据处理复杂。文献[13]设计了一种无网格算法，用于估计互质阵列中的相干信号的 DOA。然而，当相干源的相位差达到一定程度时，这种无网格算法的性能可能会下降，甚至可能无法检测到相干源。文献[14]将阵列插值和低秩矩阵恢复方法相结合，重构了增广协方差矩阵，提高了角度估计的精度和分辨率，但该方法的计算复杂度较高。此外，该方法对阵列结构的依赖性较强，当阵列模型出现偏差或数据不满足低秩假设时，估计性能可能会受到影响。

为了解决稀疏阵列 DOA 估计中目标空间谱的旁瓣增高，易受噪声影响并出现虚警现象的问题，本文提出了一种针对 TDM (time division multiplexing)

lexing) 收发模式的毫米波雷达多维目标参数处理方法，并对基于稀疏阵列原子范数最小化重构协方差矩阵 DOA 估计算法进行改进。首先通过 2D-FFT 提取出目标的距离、速度信息，然后在对接收信号矩阵进行运动补偿后进行 DOA 估计，利用奇异值阈值处理虚拟线性阵列的多个快照信号，获得预处理的协方差矩阵作为先验矩阵。然后，通过构建原子范数最小化问题，二次恢复出稀疏阵列的协方差矩阵用以精确 DOA 估计。

本文的其余部分组织如下：在第 1 章中，推导并介绍了稀疏面阵的接收信号模型。在第 2 章中，阐述了处理目标参数的算法原理，以及改进基于原子范数的稀疏阵列重构协方差矩阵 DOA 估计算法流程并进行了理论分析，然后在第 3 章中展示了仿真结果，验证了算法的有效性，最后在第 4 章得出结论并对研究做出总结。

1 信号模型

本文假设采用 TDM 收发模式的 MIMO 雷达，且它发射信号为线性调频连续波，发射信号 $s(t)$ 的表达式为：

$$s(t) = \exp\left(j2\pi\left(f_c t + \frac{1}{2}\mu t^2\right)\right) \quad (1)$$

其中， f_c 为雷达的载波频率， μ 为线性调频斜率，那么经过时延 τ 后接收到的接收信号模型为：

$$s_r(t) = \exp\left(j2\pi\left(f_c(t-\tau) + \frac{1}{2}\mu(t-\tau)^2\right) + \psi_k\right) \quad (2)$$

其中， ψ_k 为由阵元位置引入的相位差和目标运动带来的相位误差。TDM 模式下雷达不同通道发射信号和接收信号的关系如图 1 所示。

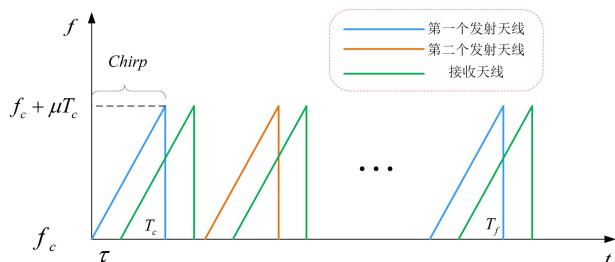


图 1 线性调频连续波接收、发射信号模型

将接收信号与发射信号进行混频以及低通滤波处理，获得的中频信号可表示为：

$$s_{IF}(t) = \exp\left(j2\pi\left(f_c\tau + \mu\tau t - \frac{1}{2}\mu\tau^2\right) + \psi_k\right) \quad (3)$$

通过对中频信号进行处理可以得到目标的多维信息，为方便信号处理将之后的数据表示为矩阵形式。若该 MIMO 雷达生成一个在方位向为均匀阵列，在俯仰向为稀疏阵列的虚拟阵列。该 $M \times N$ 个阵元的稀疏矩形阵列如图 2 所示，其在方位向相邻阵元的间距为 $d = \lambda / 2$ ，在俯仰向为相应位置 d 的整数倍。那么沿方位向排布的第一列 M 个阵元的子阵方向向量可以表示为：

$$\mathbf{a}_x(\phi, \theta) = [1, e^{-j2\pi d \cos \phi \sin \theta / \lambda}, \dots, e^{-j2\pi(M-1)d \cos \phi \sin \theta / \lambda}]^T \quad (4)$$

沿俯仰向排布的第一行 N 个阵元的子阵方向向量可以表示为：

$$\mathbf{a}_y(\phi, \theta) = [1, e^{-j2\pi d \sin \phi \sin \theta / \lambda}, \dots, e^{-j2\pi(N-1)d \sin \phi \sin \theta / \lambda}]^T \quad (5)$$

其中， ϕ 代表目标到阵列的方位角， θ 代表目标到阵列的俯仰角， λ 为波长。

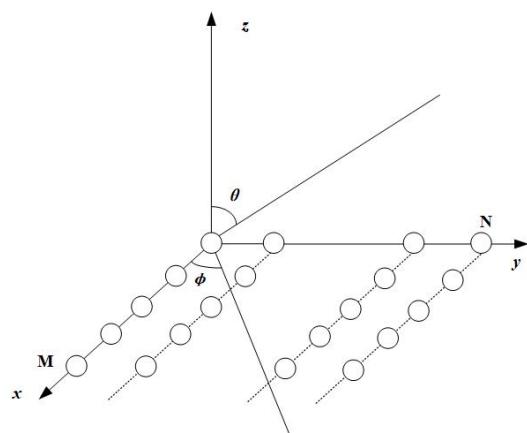


图 2 二维稀疏面阵虚拟阵列模型

现假设在远场情况下，使用该阵列发射线性调频连续波，并且接收到 K 个目标分别以方位角、俯仰角 $[(\phi_1, \theta_1), (\phi_2, \theta_2), \dots, (\phi_K, \theta_K)]$ 返回的回波信号，那么该稀疏阵列在 x 轴方向的第一个子阵的方向矩阵可表示为下式：

$$\mathbf{A}_x = [\mathbf{a}_x(\phi_1, \theta_1), \mathbf{a}_x(\phi_2, \theta_2), \dots, \mathbf{a}_x(\phi_K, \theta_K)] \quad (6)$$

那么由此子阵接收到 K 个目标的接收信号可表示为式 (7) 的矩阵形式：

$$\mathbf{x}_1(t) = \mathbf{A}_x \mathbf{s}(t) + \mathbf{n}_1(t) \quad (7)$$

其中， $\mathbf{s}(t)$ 为发射信号向量， $\mathbf{n}_1(t)$ 为该子阵的加性

噪声向量。

稀疏阵列由分布在 y 轴的 N 个子阵组成, 由此可以得到整个稀疏阵列的接收信号, 其矩阵形式表达式为:

$$\mathbf{x}(t) = \begin{bmatrix} \mathbf{x}_1(t) \\ \mathbf{x}_2(t) \\ \vdots \\ \mathbf{x}_N(t) \end{bmatrix} = \begin{bmatrix} \mathbf{A}_x \\ \mathbf{A}_x \boldsymbol{\Phi} \\ \vdots \\ \mathbf{A}_x \boldsymbol{\Phi}^{N-1} \end{bmatrix} \mathbf{s}(t) + \begin{bmatrix} \mathbf{n}_1(t) \\ \mathbf{n}_2(t) \\ \vdots \\ \mathbf{n}_N(t) \end{bmatrix} = \mathbf{As}(t) + \mathbf{n}(t) \quad (8)$$

其中:

$$\boldsymbol{\Phi} = \text{diag}\left(\left[1, e^{-j2\pi d \sin \phi_1 \sin \theta_1 / \lambda}, \dots, e^{-j2\pi d \sin \phi_K \sin \theta_K / \lambda}\right]^T\right)$$

矩阵 \mathbf{A} 表示整个面阵的方向矩阵, 其可以由下式计算:

$$\mathbf{A} = \mathbf{A}_y \odot \mathbf{A}_x = [\mathbf{a}_y(\phi_1, \theta_1) \otimes \mathbf{a}_x(\phi_1, \theta_1), \dots, \mathbf{a}_y(\phi_K, \theta_K) \otimes \mathbf{a}_x(\phi_K, \theta_K)] \quad (9)$$

其中, $\mathbf{A}_y \odot \mathbf{A}_x$ 表示 \mathbf{A}_y 和 \mathbf{A}_x 的 Khatri-Rao 积, \otimes 表示 Kronecker 积^[15]。

2 多维目标参数处理方法原理

在获得了原始雷达数据立方后, 需要进行算法处理才能获得目标的多维参数, 首先进行 2D-FFT 处理获得目标的距离以及速度信息, 之后经过 CFAR 检测获得不同距离以及速度的目标采样索引以及多普勒索引, 在运动补偿后进行 DOA 估计得到同一距离多普勒门下的目标个数以及其对应方位角和俯仰角。下面将详细介绍算法处理流程及基本原理。

2.1 测距、测速原理

首先对中频信号 $s_{\text{IF}}(t)$ 进行 2D-FFT 处理, 分别沿雷达数据立方的采样维以及脉冲维进行 DFT 处理, 获得距离-多普勒矩阵如下式:

$$\begin{aligned} DFT[s_{\text{IF}}(t)] &= \sum_{k=1}^Q \exp(j\psi_k) \cdot \exp\left(-j4\pi f_c \frac{R_0}{c}\right) \cdot \\ &\quad \sin c\left(T_p \left(f_r - 2\mu \frac{R_n}{c}\right)\right). \quad (10) \\ &\quad \sin c\left(T_a \left(f_d - \frac{2f_c v}{c}\right)\right) \end{aligned}$$

其中, R_0 为雷达发射第一个脉冲时与目标的径向距离, 即初始距离。 T_p 为扫频时间, 即一个 chirp 的持续时间, n 为一帧发射的脉冲个数, $T_a = n \cdot T_p$

是由 n 个 chirp 构成的相干处理时间, f_r 和 f_d 分别对应为中频信号频率和多普勒频率, 分别可由距离-多普勒矩阵的距离峰值索引和多普勒峰值索引获得。那么目标的距离可以通过式 (11) 得出:

$$R = \frac{c \cdot f_r}{2\mu} \quad (11)$$

目标的速度可由式(12)解算:

$$v = \frac{f_d \cdot \lambda}{2} \quad (12)$$

在经过 CFAR 检测获得 RD 图中的目标峰值, 获得目标对应的索引, 解算出目标的距离和速度信息之后, 可以在同一距离速度门下进行方位角以及俯仰角估计。

2.2 运动补偿

在进行 DOA 估计之前, 由于采用 TDM-MIMO 雷达模式, 每个发射天线在不同的时间发射信号, 目标运动会在切换发射天线产生的时延里引入额外的相位误差, 所以需要在测角之前对接收数据进行运动补偿。

现假设接收到的信号在存在多普勒频移的情况下, 其相位会产生一个变化量:

$$\Delta\phi = 2\pi f_d t \quad (13)$$

假设由式(10)得到的雷达信号矩阵第 m 行, 第 n 列元素为 $Z_{m,n}$, $0 \leq m \leq N_s$, $-N/2 \leq n \leq N/2$, 其中 N_s 为采样点数, 那每个虚拟通道补偿后的数据为:

$$\tilde{Z}_{m,n} = Z_{m,n} \cdot \exp\left(-j\frac{4\pi v(n-1)T_p}{\lambda}\right) \quad (14)$$

通过式 (14) 对目标产生的相偏值进行精确估计和相位补偿来解决速度扩展和提高测角的准确度。

2.3 测角原理

相较于均匀阵列, 实际应用中由于阵元数目的限制, 常使用稀疏面阵, 通过非均匀排布阵元来获取更大的阵列孔径以及分辨率, 但是也会导致 DOA 估计结果的旁瓣增大, 降低角度估计的准确程度。传统二维 DOA 估计算法如 2D MUSIC 算法、2D Unitary-Esprit 算法等^[16], 适用于均匀矩形阵列和 L 型阵列以及其他阵元均匀排布的阵列, 不能直接用于稀疏面阵, 需要对其进行阵元补全再进行 DOA 估计算法处理。

本文提出了一种改进的基于原子范数最小化重构协方差矩阵的二维 DOA 估计算法，首先使用改进的原子范数最小化方法通过估计协方差矩阵来实现阵元数据的填充，接着 2D MUSIC 算法估计目标方位角以及俯仰角，该算法可以精确估计目标的方位角以及俯仰角，下面介绍算法主要流程。

首先，从稀疏阵列接收信号取出俯仰阵元对应的第 i 个快拍矢量 $\mathbf{x}_i(t)$ ，通过对空缺阵元数据补零，构造阵元数为 N_{\max} 的虚拟线性阵列数据 \mathbf{Z} ，以 \mathbf{Z} 为行向量构造 Toeplitz 矩阵：

$$\mathbf{T}_v = \begin{bmatrix} \mathbf{Z}_0 & \mathbf{Z}_1^* & \cdots & \mathbf{Z}_{N_{\max}}^* \\ \mathbf{Z}_1 & \mathbf{Z}_0 & \cdots & \mathbf{Z}_{N_{\max}-1}^* \\ \vdots & \vdots & & \vdots \\ \mathbf{Z}_{N_{\max}} & \mathbf{Z}_{N_{\max}-1} & \cdots & \mathbf{Z}_0 \end{bmatrix}_{N_{\max} \times N_{\max}} \quad (15)$$

随后对 \mathbf{T}_v 进行奇异值阈值处理，获得重构的 Toeplitz 矩阵 \mathbf{T}'_v ，提取出稀疏阵列接收信号中目标的关键信息并压缩数据。下面对该 Toeplitz 矩阵进行奇异值分解：

$$\mathbf{T}_v = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^H, \boldsymbol{\Sigma} = \text{diag}(\{\lambda_i\}_{0 \leq i \leq N}) \quad (16)$$

其中， \mathbf{U} 对应为 $N \times K$ 的阵列流形矩阵， $\boldsymbol{\Sigma}$ 为对角线为奇异值的对角矩阵。根据经验与实验设定奇异值阈值为 τ ^[17]，经过奇异值分解后，对对角矩阵中的奇异值进行阈值化处理，将小于 τ 的奇异值置零后得到新对角矩阵 $\boldsymbol{\Sigma}'$ ，重构出补全数据后的 Toeplitz 矩阵：

$$\mathbf{T}'_v = \mathbf{U}\boldsymbol{\Sigma}'\mathbf{V}^H \quad (17)$$

接着，取重构 Toeplitz 矩阵 \mathbf{T}'_v 的第一行作为阵元补全后的虚拟线性阵列单快拍数据^[18]，该行向量表达式为：

$$\mathbf{Z}_v = \text{vec}[\mathbf{T}'_v] = \text{vec} \left[\sum_{k=1}^K \mathbf{a}(\theta_k) \mathbf{a}^H(\theta_k) + \sigma^2 \mathbf{I} \right] \quad (18)$$

其中， $\mathbf{a}(\theta_k)$ 为第 k 个目标的稀疏阵列接收信号导向矢量， σ^2 表示噪声方差， \mathbf{I} 为单位矩阵。生成的重构单快拍信号可看作相同孔径下，均匀线性阵列对目标采样所得的接收信号数据：

$$\mathbf{Z}_v = \mathbf{a}^*(\theta_1, \theta_2, \dots, \theta_k) p_k(t) + \sigma^2 \mathbf{I} \quad (19)$$

其中， $\mathbf{a}^*(\theta_k)$ 为虚拟线性阵列导向矢量， $p_k(t)$ 为等效均匀线阵的接收信号。由此重构了接收信号矢

量，将多快拍的数据重排构建出等效孔径的虚拟均匀线性阵列。

接下来，遍历稀疏阵列接收信号的所有快拍，按快拍顺序将对应快拍矢量 \mathbf{Z}_v 补全为虚拟线性阵列协方差矩阵，重构出的协方差矩阵 \mathbf{R}_v 如下：

$$\mathbf{R}_v = [\mathbf{Z}_{v1}, \mathbf{Z}_{v2}, \dots, \mathbf{Z}_{vn}]_{N_{\max} \times L} \quad (20)$$

其中， \mathbf{Z}_{vn} 代表由第 n 个快拍恢复的 $1 \times N_{\max}$ 维单快拍数据。以该恢复出的虚拟线性阵列协方差矩阵 \mathbf{R}_v 作为先验，构建原子范数最小化问题^[19]求解用以 DOA 估计的协方差矩阵。首先确定观测角度的连续原子集：

$$A_r = \{r(\theta) | \theta \in [-90^\circ, 90^\circ]\} \quad (21)$$

该观测角度下的原子范数如下式^[20]：

$$\|z\|_{A_r} = \inf\{K : z = \sum_{k=1}^K \mathbf{a}(\theta_k) s_k, \theta_k \in [-90^\circ, 90^\circ]\} \quad (22)$$

其中， $\|z\|_{A_r}$ 表示在观测角度集合 A_r 约束下的原子范数， K 为原子数量， s_k 表示基函数 $\mathbf{a}(\theta_k)$ 的相应权重。之后可通过求解如下原子范数最小化问题(23)来获得想要恢复的稀疏协方差矩阵：

$$\begin{aligned} \min \frac{1}{2} \|\mathbf{T}(z) - \mathbf{R}_v\|_F^2 + \mu \text{tr}(\mathbf{T}(z)) &\leq \varepsilon \\ \text{s.t. } \mathbf{T}(z) &\geq 0 \end{aligned} \quad (23)$$

通过遍历所有的快拍数，取每次迭代原子范数计算所得的向量作为新的快拍接收信号，组合成新的协方差矩阵。由此可将俯仰向稀疏的阵元补全，生成均匀面阵等效数据。式(23)中的重构协方差矩阵 $\mathbf{T}(z)$ 可视为稀疏阵列相同孔径下虚拟均匀矩形阵列的协方差矩阵，再使用传统均匀线性阵列 DOA 估计方法进行 DOA 估计。本文选择使用 MUSIC 算法进行 DOA 估计，MUSIC 算法求解方位角功率谱表达式如下：

$$P_{\text{MUSIC}} = \frac{1}{\mathbf{a}^H(\theta) \mathbf{U}_n \mathbf{U}_n^H \mathbf{a}(\theta)}, \theta \in [-90^\circ, 90^\circ] \quad (24)$$

其中， $\mathbf{a}(\theta)$ 为虚拟均匀线性阵列的导向矢量， \mathbf{U}_n 为重构协方差矩阵的噪声子空间。

上述过程为本文提出改进基于原子范数的稀疏阵列重构协方差矩阵 DOA 估计算法的具体步骤，算法的简要流程图如图 3 所示。

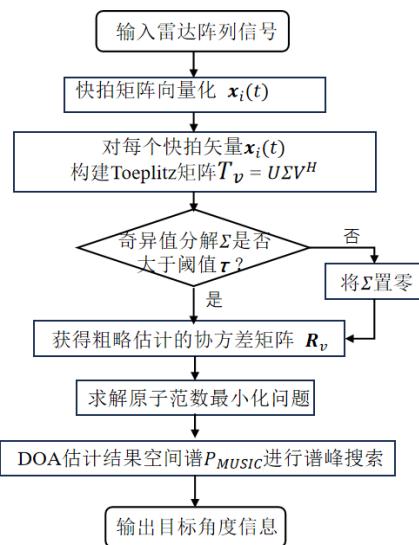


图 3 DOA 估计算法流程图

3 仿真实验

为了分析本文所提出改进算法的 DOA 估计性能、稳定性以及处理速度,进行了以下几组仿真实验,通过对比该算法与其他算法的处理效果,验证了本文所提算法的有效性。仿真实验皆在操作系统为 Windows11, 处理器为 12 代 Intel i7-12700H, 内存 16 GB 的 64 位笔记本上使用 MATLAB 软件进行仿真, 算法中求解原子范数最小化重构协方差矩阵问题使用 CVX 工具箱求解。

实验参数的具体配置如下: 仿真实验使用均匀线性阵列作为方位向的接收天线, 采用最小冗余阵列作为俯仰向的接收天线构成一个稀疏面阵。其中俯仰向的阵元数量为 7, 阵元物理位置为 $[0,1d,2d,6d,10d,14d,17d]$, d 为阵元间距等于半波长, 快拍数为 100。在不同算法进行 DOA 估计性能比较实验中进行了 100 次蒙特卡洛实验, 并取均值作为实验结果。根据实验估计值与设置的角度真值所计算的均方根误差作为评价指标, 其中均方根误差计算公式如下:

$$RMSE = \sqrt{\frac{1}{KQ} \sum_{k=1}^K \sum_{q=1}^Q [(\hat{\theta}_k - \theta_k)^2 + (\hat{\varphi}_k - \varphi_k)^2]} \quad (25)$$

其中, $\hat{\theta}_k$ 和 $\hat{\varphi}_k$ 为仿真实验设定的目标方位角和俯仰角真值, θ_k 和 φ_k 为 DOA 估计估算的方位角和俯仰角, K 为目标个数, Q 为蒙特卡洛实验次数。实验具体的仿真参数如表 1。

表 1 仿真参数

仿真参数	设置参数值
Chirp 数	512
采样点数	256
起始频率/GHz	76.5
接收天线位置	$[0,1d,2d,6d,10d,14d,17d]$
Chirp 斜率 μ	$3E+13$
扫频时间/ μ s	10
帧数 n	100

3.1 算法的分辨性能仿真实验

在 DOA 估计中, 分辨力是指阵列处理算法能够区分来自不同方向的信号的能力。它描述了算法在角度空间中能够分辨两个相邻信号源的最小角度差。为检验该算法的分辨性能, 该实验设置了两个点目标, 与雷达的径向距离、径向速度、方位角和俯仰角分别设置为 $[10 \text{ m}, -5 \text{ m/s}, -10^\circ, -15^\circ]$, $[15 \text{ m}, 3 \text{ m/s}, -15^\circ, -20^\circ]$, 首先进行 2D-FFT 处理接收信号矩阵生成目标的距离-多普勒图如图 4 所示。

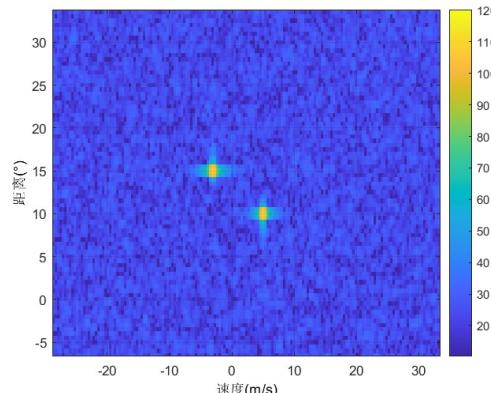


图 4 距离-多普勒图

对图 4 获得的 RD 矩阵进行运动补偿后, 再对俯仰向的稀疏阵元进行阵元补全, 生成等效的均匀矩形阵列进行 2D-MUSIC 处理, 生成的方位角-俯仰角空间谱图如图 5 所示。可以在空间谱中区分出该两个相近的点目标。

在图 5 的方位角-俯仰角空间谱图中, 可以清晰地观察到在仿真设定真值的点目标位置处出现了显著的峰值。这表明通过对 RD 矩阵进行运动补偿并采用改进原子范数最小化的协方差矩阵重构 DOA 估计算法处理实现了阵元补全以及 DOA 估计, 成功构建了等效的均匀矩形阵列, 从而显著提升了 2D-MUSIC 算法的分辨能力。在图 5 的空间

谱中, 尽管目标之间的角度差异较小, 但依然能够在空间谱图中明显区分出这两个点目标, 验证了该方法在多目标 DOA 估计中的有效性。这一结果进一步说明了通过稀疏阵列的阵元补全技术可以有效提高 DOA 估计中的分辨力, 为复杂场景下的多目标定位。

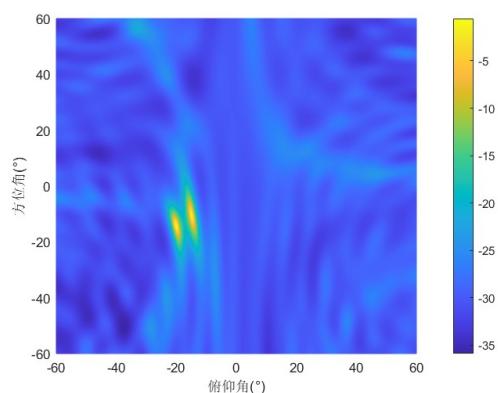


图 5 方位角-俯仰角空间谱

3.2 不同 DOA 算法性能对比仿真实验

在不同 DOA 算法性能对比仿真实验中, 对所提出的基于原子范数最小化的协方差矩阵重构改进算法以及其他三种算法(空间平滑算法(SS)、核范数最小化算法(SS)、原子范数最小化算法(ANM))进行了比较。分别在不同信噪比(SNR)和不同快拍数下进行了两组实验, 并计算算法结果的均方根误差作为评价指标。其中设置信噪比范围为-30 dB 至 30 dB, 目标个数为 6 个, 俯仰角分别为 $[-60^\circ, -35^\circ, -15^\circ, 15^\circ, 35^\circ, 60^\circ]$, 快拍数范围为 50 至 500, 进行了 100 次蒙特卡罗模拟实验。实验结果如图 6 和图 7 所示。

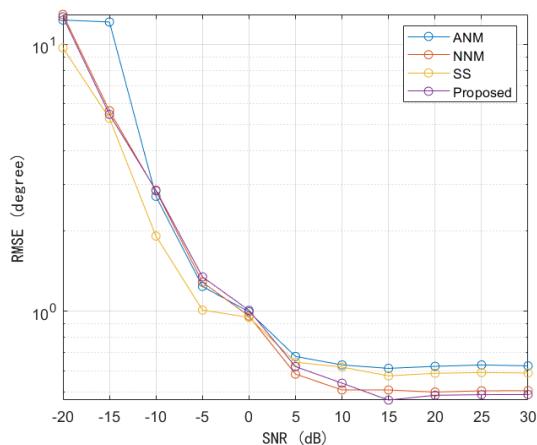


图 6 不同信噪比下 DOA 估计算法性能对比

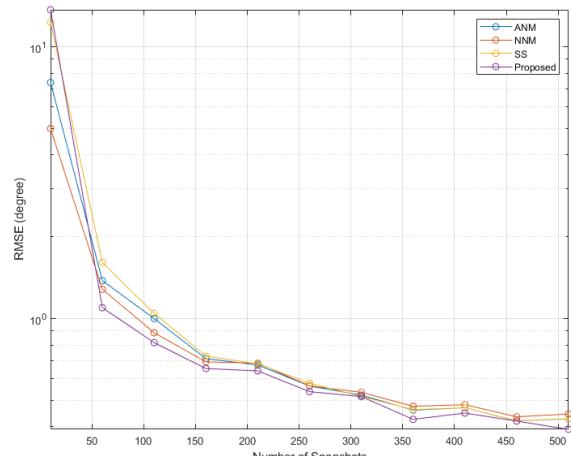


图 7 不同快拍数下 DOA 估计算法性能对比

从图 6 和图 7 可以得出, 随着信噪比(SNR)的提高和快拍数的增加, 算法估计结果与真值的均方根误差逐渐减小。在多目标场景下, 与其他方法相比, 空间平滑 SS-MUSIC 方法在低信噪比条件下表现出较大的均方根误差^[20]。所提出算法的 DOA 估计均方根误差在相似条件下与 NNM 和 ANM 方法接近。当信噪比大于 0 dB 时, 所提出的算法优于其他算法。基于本次实验结果, 可以得出结论: 所提出的算法实现了高性能的 DOA 估计, 对稀疏阵元数据的补全, 改进的基于原子范数最小化的协方差矩阵重构 DOA 估计算法性能随信噪比的改善而增强。

4 结语

本文针对稀疏面阵提出了一种基于毫米波雷达多维目标参数处理方法, 并对基于原子范数最小化的协方差矩阵重构 DOA 估计算法进行改进。该处理方法通过 2D-FFT 提取目标的距离、速度信息, 在运动补偿后进行高精度 DOA 估计出目标的方位角与俯仰角, 确定目标的空间位置信息以及运动参数信息。改进算法首先利用奇异值阈值处理构建虚拟线性阵列多快拍信号作为粗略估计, 再通过构造原子范数最小化问题对稀疏协方差矩阵进行重构, 恢复出精准协方差矩阵进行 DOA 估计。最后通过仿真实验证了该算法的有效性。

(下转第 199 页)

基于俯仰角反演的雷达重定位误差补偿方法研究

王彦平，崔彭，白泽朝，申文杰，林贊，李洋

北方工业大学 信息学院，北京 100144

摘要：地基合成孔径雷达在非连续监测模式时会产生重定位误差，严重影响形变测量精度。现有方法主要基于雷达与目标几何关系建立多参数模型，通过最小二乘法估计参数对其进行补偿。但参数估计需要俯仰角信息，实际应用中存在限制。为此提出一种基于俯仰角反演的雷达重定位误差补偿方法，对雷达垂直方向变化产生的相位进行分析，建立相位与俯仰角之间关系，通过相位变化获取目标俯仰角信息；将计算得到的俯仰角结合多参数模型，利用最小二乘法对误差进行补偿；通过仿真实验进行了分析。实验结果表明：俯仰角正余弦反演精度在 0.1 以内，补偿后残差相位均方根误差为 0.0013 弧度，较传统线性三角模型降低了 2.2999 弧度。该方法无需外部辅助即可获取目标俯仰角信息，尤其是在大俯仰角场景能够实现高精度位置误差补偿。

关键词：GBSAR；重新定位误差补偿；多参数模型；俯仰角反演；非连续监测模式

文献标志码：A

中图分类号：TP391

Research on Radar Repositioning Error Compensation Method Based on Pitch Angle Inversion

WANG Yanping, CUI Peng, BAI Zechao, SHEN Wenjie, LIN Yun, LI Yang

School of Information Science and Technology, North China University of Technology, Beijing 100144, China

Abstract: Repositioning errors in ground-based synthetic aperture radar (GBSAR) systems operating in discontinuous monitoring modes can significantly degrade the accuracy of deformation measurements. Current methods for error compensation primarily rely on multi-parameter models that are constructed using the geometric relationship between the radar and the target, with parameters estimated through least squares techniques. However, these methods often require pitch angle data, which can be challenging to obtain in real-world applications. To overcome this issue, this paper introduces a radar repositioning error compensation technique based on pitch angle inversion. The proposed approach begins by analyzing the phase shifts caused by vertical displacements of the radar, establishing a mathematical relationship between the phase and the pitch angle. The pitch angle is then extracted from the phase data and integrated into a multi-parameter model for error correction using the least squares method. Simulations show that the proposed method achieves pitch angle inversion with an accuracy of within 0.1, and the root mean square error of the corrected phase residuals is 0.0013 radians, which is 2.2999 radians lower than that of conventional linear triangular models. This approach does not require external assistance to obtain pitch angle information and is particularly effective in scenarios with large pitch angles for high-precision error correction.

Key words: GBSAR; repositioning error; multi-parameter model; pitch angle inversion; discontinuous measurement

基金项目：国家重点研发青年科学家项目（2023YFB3905200）；北京市教育委员会科学研究计划项目（KM202410009001）。

作者简介：王彦平（1976—），男，博士，研究员，研究方向为雷达三维成像、地基 SAR 系统研制及其应用；崔彭（2000—），男，硕士研究生，研究方向为地基 SAR 重新定位误差校正；白泽朝（1991—），男，博士，研究方向为星载/地基 InSAR 技术理论和应用研究。

地基合成孔径雷达 (ground-based synthetic aperture radar, GBSAR) 是一种高精度地表形变监测设备, 具备全天候、高精度及近实时的形变测量能力, 被广泛应用于滑坡^[1-5]和冰山融化^[6-7]等形变测量中。一般 GBSAR 形变监测模式包括连续性监测和非连续性监测。其中连续监测模式下, 设备需要长时间固定安装, 每隔几分钟对场景进行一次扫描成像^[8]。然而, 连续监测容易导致数据冗余, 同时存在较高的运维成本。非连续监测模式下, 设备则无需长时间固定, 可灵活调整观测周期, 间歇扫描。因此对于缓慢形变的测量场景, 非连续监测模式更适用且成本更低^[9]。

然而, 非连续监测模式下, 雷达设备反复安装和拆卸, 不可避免地导致雷达位置偏移, 产生重定位误差, 严重影响形变监测精度^[10]。针对非连续监测模式的重定位误差补偿, 已有学者开展了相关内容的研究。Hu 等人^[11]基于三元函数的泰勒展开对重定位误差进行详细的理论分析, 并建立了线性多参数模型以进行补偿。Yang 等人^[12]提出了一种新的重定位误差修正方法, 将重定位误差分解为距离、方位和垂直方向的三个分量, 并通过最小二乘法求解补偿, 为非连续监测模式在工程中的应用提供了可能性。Mo 等人^[13]提出了基于泰勒展开的基线误差建模数学框架。在此基础上, 提出了一种新的多参数非线性三角模型, 用于精确补偿基线误差相位。为了提高补偿效率, 还构建了一种高效的线性三角模型。然而, 该模型进行最小二乘法参数估计时, 需要借助外部辅助获取目标俯仰角信息, 限制了其应用范围。

针对此问题, 本文提出了一种基于俯仰角反演的雷达重定位误差补偿方法。首先分析雷达与目标几何关系, 通过干涉相位获取雷达与目标之间俯仰角信息。并基于所得俯仰角信息, 利用多参数非线性模型与最小二乘法参数估计实现对重定位误差的补偿。该方法可在大俯仰角场景下准确补偿重定位误差, 无需额外辅助信息, 扩宽了应用的场景。

1 重新定位误差补偿方法分析

1.1 基于多参数非线性模型的重新定位误差补偿方法分析

GBSAR 两次观测获取的干涉相位可以表示

为^[14]:

$$\Delta\varphi = \varphi_2 - \varphi_1 = \frac{4\pi}{\lambda}(\rho_2 - \rho_1) = \frac{4\pi}{\lambda}\Delta\rho \quad (1)$$

其中, ρ_2 和 ρ_1 分别是 GBSAR 两次观测时目标点与雷达距离, λ 为波长, $\Delta\varphi$ 为干涉相位。

在非连续监测模式中, 雷达多次装卸导致位置偏移, 进而造成额外的距离变化, 即雷达与目标之间的电磁波传播路径发生变化, 在回波中产生相位误差, 其不仅包括目标点形变引起的距离变化, 也包括雷达位置移动造成的位置变化, 因此在干涉相位中引入了重定位误差相位, 使得 $\Delta\varphi$ 不能准确反映形变^[15]。其可根据雷达与目标的几何关系建立数学模型, 建立重定位误差相位与雷达位置之间的关系, 通过最小二乘法估计参数以进行补偿。

如图 1 所示, 将雷达位置误差分为沿坐标轴的分量。以沿 x 轴的雷达误差分量为例进行分析, 观测几何图形可简化为图 1 (a) 所示。目标点, 雷达原始位置, 雷达偏移后位置形成三角形, 此时可以根据余弦定理将两次斜距之差用俯仰角 α 、方位角 β 及雷达 x 轴误差分量 B_x 表示。

$$OP - OP' = \sqrt{\rho^2 + B_x^2 + 2 \cdot \rho \cdot B_x \cdot \cos\beta \cos\alpha} \quad (2)$$

对其泰勒级数展开, 忽略二次项, 保留一次项可得:

$$OP - OP' = \cos\beta \cos\alpha B_x \quad (3)$$

将雷达 x 轴方向上重定位误差相位分量表示为:

$$\Delta\varphi_x = \frac{4\pi}{\lambda}(B_x \cos(\beta) \cos(\alpha)) \quad (4)$$

同理对 y 轴和 z 轴方向上误差分量分析, 可将雷达重新定位误差相位表示为三个分量之和, 结果如下:

$$\begin{aligned} \Delta\varphi_{\text{repo}} &= \Delta\varphi_x + \Delta\varphi_y + \Delta\varphi_z = \\ &\frac{4\pi}{\lambda}(B_x \cos(\beta) \cos(\alpha) + B_y \sin(\beta) \cos(\alpha) + B_z \sin(\alpha)) = \\ &a_1 + a_2 \cos(\beta) \cos(\alpha) + a_3 \sin(\beta) \cos(\alpha) + a_4 \sin(\alpha) \end{aligned} \quad (5)$$

a_1 、 a_2 、 a_3 和 a_4 为待估计参数, a_1 为模型偏差, a_2 、 a_3 、 a_4 与三个方向上的空间基线分量有关。此时, 利用合适的选点策略^[16-18]从观测场景中提取永久散射体, 通过式 (5), 采用最小二乘法对参数进行估计, 即可补偿重定位误差相位。

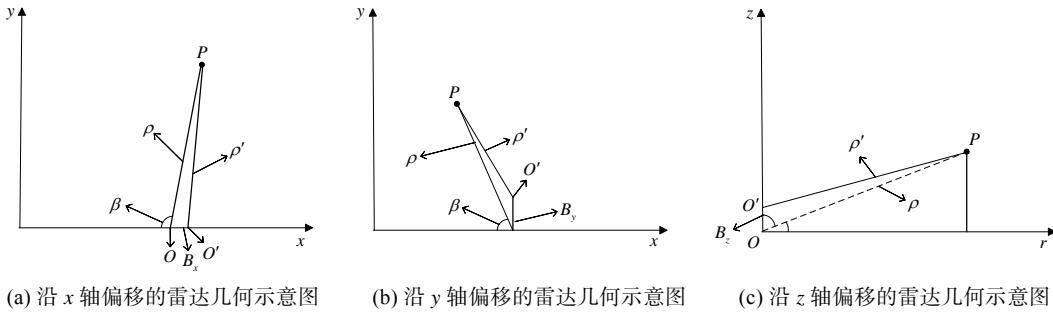


图 1 雷达偏移几何示意图

在参数估计时,需要目标相对于雷达的方位信息和俯仰角信息。采集数据时可以很容易地获取目标斜距和方位信息,但俯仰角信息的获取比较困难。因此,面对俯仰角范围较大的场景,本文基于雷达高度变化获取俯仰角信息。

1.2 基于雷达高度变化的俯仰角反演

一般情况下,俯仰角可以通过额外的数字高程模型或全站仪获取,但会增加形变监测的复杂度,降低实用性。因此,本文基于雷达与目标几何关系对干涉相位建模,建立相位与俯仰角信息关系,在不依赖数字高程模型和全站仪的情况下,可精确获取俯仰角。

其基本原理为通过调整雷达高度,提取 z 轴方向上雷达位置变化引起的干涉相位,利用相位变化计算出所需的俯仰角信息。

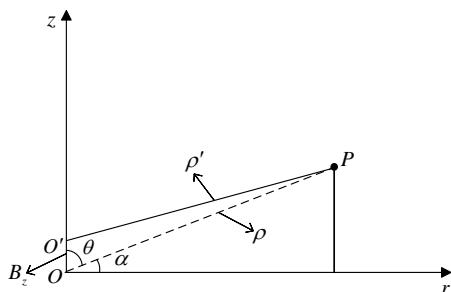


图 2 GBSAR 图像采集几何图

图 2 为雷达高度改变与目标的几何关系。 O 为雷达原始位置, O' 为雷达改变高度后位置,雷达 z 轴方向上偏移量用 B_z 表示。两次观测间隔较短,设 P 点未发生形变,根据余弦定理推出:

$$|O'P|^2 = |OO'|^2 + |OP|^2 - 2 \cos \theta |OO'| |OP| \quad (6)$$

式中 $|\cdot|$ 为欧式距离算子,改写如下:

$$\rho - \rho' = -\frac{B_z^2 - 2 \cos \theta \cdot \rho \cdot B_z}{\rho' + \rho} \quad (7)$$

形变监测时,雷达与观测目标之间的距离通常在几百米到几千米左右,雷达的位置变化可控制在毫米到分米范围内,所以 $B_z \ll \rho$,且 $\rho' + \rho \approx 2\rho$,可得出下式:

$$d_z = \rho - \rho' \approx -\frac{B_z^2}{2\rho} + B_z \cos \theta \approx B_z \cos \theta \quad (8)$$

由图 2 知 $\theta = \frac{\pi}{2} - \alpha$, 得出下式:

$$d_z = B_z \sin \alpha \quad (9)$$

综上,得出雷达高度变化时所产生干涉相位表达式如下:

$$\Delta \varphi_{repo} = \frac{4\pi}{\lambda} B_z \sin \alpha \quad (10)$$

GBSAR 干涉相位不仅包含目标形变相位,还包括一系列误差相位。相位信息可写成如下所示线性组合。

$$\Delta \varphi = \varphi_{defo} + \varphi_{atmo} + \varphi_{repo} + \varphi_{noise} \quad (11)$$

φ_{defo} 为形变相位, φ_{atmo} 为大气干扰相位, φ_{repo} 即重定位误差相位, φ_{noise} 为观测噪声。

实际应用中,雷达位置变化前后观测间隔较短,目标点的形变相位和大气干扰相位可忽略不计。通过自适应滤波对地基干涉雷达图像降噪后得到:

$$\Delta \varphi = \Delta \varphi_{repo} = \frac{4\pi}{\lambda} B_z \sin \alpha \quad (12)$$

此时干涉相位即为重定位误差相位,通过已知信息,波长 λ 与 z 轴方向上偏移量 B_z 可求出俯仰信息 $\sin \alpha$ 。

如图 3 所示,雷达改变位置前后,俯仰角变化为 $\Delta \alpha$,观测时目标点距离雷达几百米到千米,雷达偏移量可控制在毫米范围内, $\Delta \alpha$ 近似为零,可得下式:

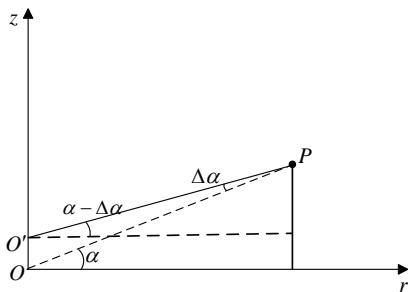


图 3 倾仰角变化几何示意图

$$\begin{aligned} \sin \alpha - \sin(\alpha - \Delta\alpha) = \\ 2 \cos \frac{2\alpha - \Delta\alpha}{2} \sin \frac{\Delta\alpha}{2} \approx 0 \end{aligned} \quad (13)$$

因此,雷达在拆装过程中虽造成雷达位置偏移,但视作目标相对于雷达的倾仰角并未发生变化。

2 仿真分析

本文方法基本流程如图 4 所示,首先,改变雷达高程,由干涉相位获取倾仰角信息。然后,使雷达水平方向和垂直方向上发生偏移,利用方位角和所得倾仰角,结合多参数非线性模型,通过最小二乘法对参数进行估计,补偿重定位误差相位。

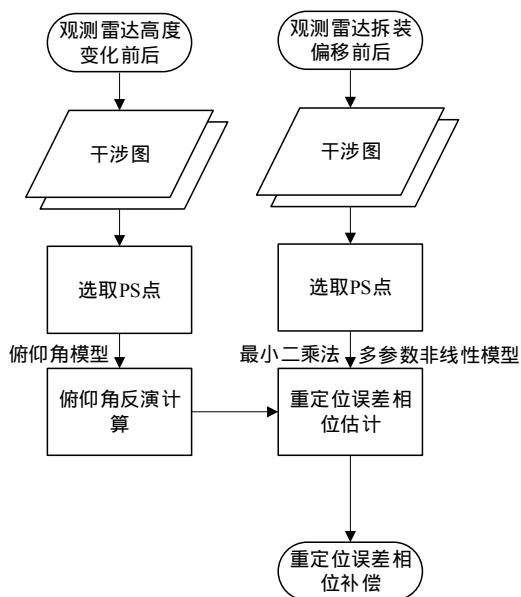


图 4 本文重定位误差补偿流程

2.1 倾仰角反演仿真

对倾仰角进行仿真,模拟高度从 15 米到 35 米的斜坡地形,雷达观测范围为 45 度到 135 度。雷达 z 轴偏移量为 2 毫米,提取干涉相位对倾仰角进行计算,得到的结果如图 5 所示。

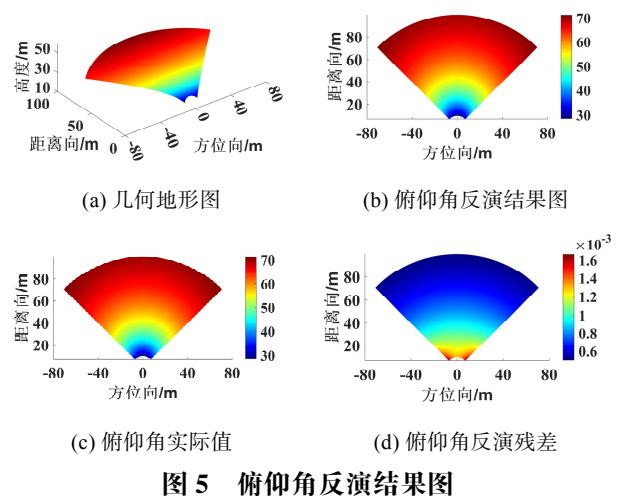


图 5 倾仰角反演结果图

图 5 (b) 为雷达 z 轴偏移量为 2 毫米情况下,提取对应干涉相位反演计算所得目标倾仰角结果图。图 5 (c) 为倾仰角实际值。将二者相减后,得到图 5 (d) 倾仰角反演残差。对倾仰角反演残差进行定量分析,其均方根误差为 8.8417×10^{-4} 度,最大误差为 0.0017 度。若要实现亚毫米级的重新定位误差补偿,倾仰角正余弦精度必须控制在 0.1 以内,本文反演的倾仰角精度满足后续误差补偿要求。

2.2 重新定位误差相位补偿仿真

基于反演倾仰角信息,利用多参数非线性模型和最小二乘法对参数进行估计,补偿重定位误差相位,其结果如图 6 所示。雷达在 x 轴、 y 轴和 z 轴方向上的偏移量为 10 毫米,产生的重新定位误差相位如图 6 (b) 所示。基于反演所得倾仰角对重定位误差相位进行补偿,结果如图 6 (c) 所示。补偿后重定位误差的残差相位如图 6 (d) 所示。其均方根误差为 0.0013 弧度,满足形变监测精度要求。

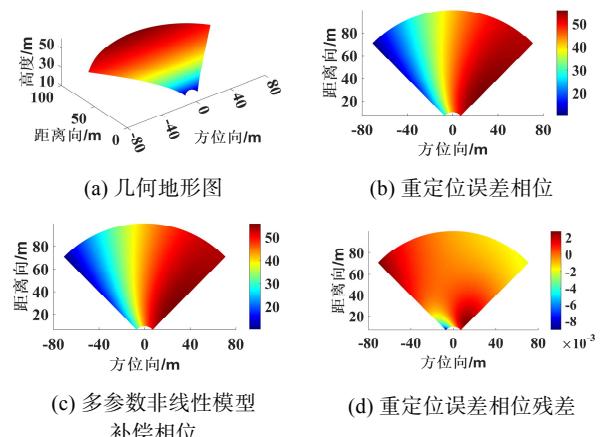


图 6 重定位误差相位补偿结果

设观测场景俯仰角 $\alpha \in [\eta, \xi]$ ，取 $\alpha^* = ((\xi - \eta)/2)$ ，带入多参数非线性模型中得 $\Delta\varphi_{\text{repo}} = a_1 + a_2 \cos(\beta) + a_3 \sin(\beta)$ ，即线性三角模型。对重定位误差相位补偿，补偿残差图如下所示。对结果进行定量分析，基于反演俯仰角对重定位误差相位进行补偿，其残差相位均方根误差为 0.0013 弧度。基于线性三角模型对重定位误差补偿，残差相位均方根误差为 2.3012 弧度。由此验证了基于俯仰角反演的雷达重定位误差补偿方法在大俯仰场景下对重定位误差补偿的有效性。

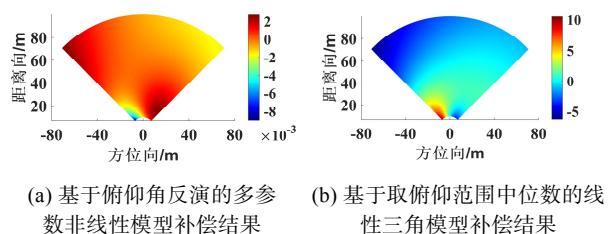


图 7 误差补偿结果对比

3 结论

本文提出了一种俯仰角反演的雷达重定位误差补偿方法。该方法通过干涉相位反演计算雷达与目标之间的俯仰角信息，利用多参数非线性模型和最小二乘法对重定位误差相位进行补偿。本文方法无需额外的外部信息，便能在大俯仰角场景下对重定位误差进行补偿。与传统的线性三角模型相比，本文方法补偿效果明显优于线性三角模型，通过仿真分析验证了该方法的有效性。

参考文献：

- [1] 白泽朝, 王彦平, 王振海, 等. 地基大视场 SAR 形变监测的非均匀大气相位校正方法[J]. 雷达学报, 2023, 12(1): 53-63.
- [2] IGLESIAS R, AGUASCA A, FÀBREGAS X, et al. Ground-based polarimetric SAR interferometry for the monitoring of terrain displacement phenomena—part II: applications[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2015, 8: 994-1007.
- [3] IGLESIAS R, AGUASCA A, FABREGAS X, et al. Ground-based polarimetric SAR interferometry for the monitoring of terrain displacement phenomena—part I: theoretical description[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2014, 8(3): 980-993.
- [4] NOFERINI L, PIERACCINI M, MECATTI D, et al. Using GB-SAR technique to monitor slow moving landslide[J]. Engineering Geology, 2007, 95(3/4): 88-98.
- [5] LEVA D, NICO G, TARCHI D, et al. Temporal analysis of a landslide by means of a ground-based SAR interferometer[J]. IEEE Transactions on Geoscience and Remote Sensing, 2003, 41(4): 745-752.
- [6] YITAYEW T G, FERRO-FAMIL L, ELTOFT T, et al. Tomographic imaging of Fjord ice using a very high resolution ground-based SAR system[J]. IEEE Transactions on Geoscience and Remote Sensing, 2016, 55(2): 698-714.
- [7] YITAYEW T G, FERRO-FAMIL L, ELTOFT T, et al. Lake and fjord ice imaging using a multifrequency ground-based tomographic SAR system[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2017, 10(10): 4457-4468.
- [8] PIERACCINI M, MICCINESI L. Ground-based radar interferometry: a bibliographic review[J]. Remote Sensing, 2019, 11(9): 1029.
- [9] MONSERRAT O, CROSETTO M, LUZI G. A review of ground-based SAR interferometry for deformation measurement[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2014, 93: 40-48.
- [10] WANG Z, LI Z, MILLS J. Modelling of instrument re-positioning errors in discontinuous multi-campaign ground-based SAR (MC-GBSAR) deformation monitoring[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2019, 157: 26-40.
- [11] HU C, ZHU J X, DENG Y K, et al. Repositioning error compensation in discontinuous ground-based SAR monitoring[J]. Remote Sensing, 2021, 13(13): 2461-2461.
- [12] YANG H, CAI J, PENG J, et al. A correcting method about GB-SAR rail displacement[J]. International Journal of Remote Sensing, 2017, 38(6): 1483-1493.
- [13] MO Y, LAI T, WANG Q, et al. Modeling and compensation for repositioning error in discontinuous GBSAR monitoring[J]. IEEE Geoscience and Remote Sensing Letters, 2023.
- [14] WANG P, XING C, PAN X. Reservoir dam surface deformation monitoring by differential GB-InSAR based on image subsets[J]. Sensors, 2020, 20(2): 396.
- [15] HU C, ZHU M, ZENG T, et al. High-precision deformation monitoring algorithm for GBSAR system: rail determination phase error compensation[J]. Science China Information Sciences, 2016, 59: 1-16.
- [16] FERRETTI A, PRATI C, ROCCA F. Permanent scatterers in SAR interferometry[J]. IEEE Transactions on Geoscience and Remote Sensing, 2001, 39(1): 8-20.
- [17] FERRETTI A, PRATI C, ROCCA F. Nonlinear subsidence rate estimation using permanent scatterers in differential SAR interferometry[J]. IEEE Transactions on Geoscience and Remote Sensing, 2000, 38(5): 2202-2212.
- [18] ISHITSUKA K, MATSUOKA T, TAMURA M. Persistent scatterer selection incorporating polarimetric SAR interferograms based on maximum likelihood theory[J]. IEEE Transactions on Geoscience and Remote Sensing, 2016, 55(1): 38-50.

基于双平行线阵和联合 IAA-RIT 的毫米波雷达高精度 2D-DOA 估计方法

于丹阳，杜 磊

中国计量科学研究院 力学与声学计量科学研究所，北京 100029

摘要：高精度二维波达角（2D-DOA）估计是毫米波雷达进行精准目标定位的重要手段。针对毫米波雷达目标探测中天线孔径受限、信号相干、少快拍等问题，提出了一种基于双平行线阵和联合迭代自适应算法与旋转不变技术（IAA-RIT）的2D-DOA估计方法。该方法基于双平行线阵中的子阵1构建迭代自适应算法谱，通过谱峰搜索获得目标水平角与俯仰角的耦合角度估计值，基于两个子阵构建扩展协方差矩阵并利用空间平滑技术进行信号解相干，基于耦合角度估计值对信号方向矩阵进行重构并计算两个子阵间的旋转不变关系，可获得另一组耦合角度估计值，基于两组耦合角度估计值并利用空间几何关系对水平角和俯仰角进行解耦合。实验结果表明，IAA-RIT 可在信噪比良好时仅使用极少快拍数据对相干信号进行高精度2D-DOA 估计，且无需进行额外的角度匹配。

关键词：毫米波雷达；迭代自适应；旋转不变技术；双平行线阵；少快拍；信号相干；角度匹配

文献标志码：A

中图分类号：TN959.6; TN98

High Precision 2D-DOA Estimation Method for Millimeter Wave Radar Based on Double-Parallel Linear Array and Joint IAA-RIT

YU Danyang, DU Lei

Division of Mechanics and Acoustics Metrology, National Institute of Metrology, Beijing 100029, China

Abstract: High precision two-dimensional direction of arrival (2D-DOA) estimation is an important mean of millimeter wave radar for accurate target location. Aiming at the problems such as limited antenna aperture, signal coherence and a few snapshots in millimeter wave radar target detection, this paper proposes a 2D-DOA estimation method by using a joint iterative adaptive approach and rotational invariance technique (IAA-RIT) based on double-parallel linear array. The method first constructs an iterative adaptive approach spectrum based on the subarray 1 in the double-parallel linear array, and then calculates the coupling angle estimates with the azimuth and the elevation. Secondly, the extended covariance matrix based on two subarrays is constructed, and the signal is decorrelated by the spatial smoothing technique. Then the signal direction matrix is reconstructed based on the coupling angle estimate and the rotational invariance relationship between the two subarrays is calculated to obtain another set of coupling angle estimates. Finally, the azimuth and the elevation are decoupled based on two sets of estimated coupling angles and the spatial geometry relation. The experimental results show that IAA-RIT can estimate the coherent signal with high precision 2D-DOA with a few snapshots and no additional angle matching.

Key words: millimeter wave radar; iterative adaptive approach; rotational invariance technique; double parallel line array; a few snapshots; signal coherence; angle matching

基金项目：国家重点研发计划（2022YFF0604803, 2017YFF0205006）；国家自然科学基金（62131001）。

作者简介：于丹阳（1999—），女，硕士研究生，研究方向为雷达目标参数校准；杜磊（1982—），男，通信作者，博士，副研究员，研究方向为运动学参数计量，E-mail：dulei@nim.ac.cn。

近年来,毫米波雷达以其体积小、成本低、全天候工作、环境适应性强等优点广泛应用于智能汽车和智能交通领域^[1]。汽车毫米波雷达通过测量目标的距离、速度和角度来获得目标的存在和位置。出于自动驾驶安全性考虑,汽车毫米波雷达需要获得更加准确的道路情况及周围环境信息^[2],高分辨率四维(距离-速度-方位角-俯仰角)成像汽车毫米波雷达成为新兴研究热点^[3],对目标进行高精度二维波达角估计(two-dimensional direction of arrival, 2D-DOA)的需求日益增长。

目前,为了在阵列孔径受限的条件下提高波达角估计的精度与分辨力,汽车毫米波雷达常使用多输入多输出(multiple input multiple output, MIMO)体制^[4-5],通过合成虚拟等效阵列来扩大天线阵列孔径。实现MIMO体制最简单直接的方法是利用时分复用(time division multiplex, TDM)技术^[6]。TDM技术通过使用不同的发射天线依次发射线性调频连续波(linear frequency modulation continuous wave, LFMCW)来实现波形正交,可轻松在接收天线处实现通道分离^[7],获得更大的等效虚拟阵列,从而实现更高精度的DOA估计。

另外,通过信号处理算法来提高DOA估计的分辨力与精度也是不容忽视的研究内容。目前,较为传统的DOA估计算法^[8-9]包括快速傅里叶变换(fast Fourier transform, FFT)、常规波束形成(conventional beam forming, CBF)、最小方差无失真响应(minimum variance distortionless response, MVDR)等。尽管这些方法原理简单,容易实现,但它们的分辨力由于受到瑞利限的影响,难以应用于一些对分辨力要求较高的场景。目前,许多学者提出了可以突破瑞利限的DOA超分辨算法。基于子空间分解的MUSIC算法与ESPRIT算法,利用噪声子空间与目标方向向量的正交性来获得目标角度,但其需要多快拍数据且在信号相干时为获得准确的DOA估计将会损失阵列孔径^[10-11]。由于雷达回波信号具有空域稀疏性,因此压缩感知理论可被应用于DOA估计^[12],如正交匹配追踪算法(orthogonal matching pursuit, OMP)、稀疏贝叶斯学习(sparse Bayesian learning, SBL)。李子高等人^[13]提出了一种局部优化处理的OMP方法,减少了算法的计算复杂度,但其角度分辨力有待提

高。张宇等人^[14]采用期望最大化算法,对稀疏贝叶斯模型的离散网格节点进行迭代修正,基于稀疏贝叶斯学习实现了一种离网格DOA估计算法,同时数值实验证明了该算法在低信噪比、少快拍数的测量条件下能获得较好的DOA估计结果,但其计算复杂度较高。

迭代自适应算法(iterative adaptive approach, IAA)通过迭代信号协方差矩阵与能量矩阵进行高精度DOA估计。不仅能够在信号相干和少快拍情况下工作良好,且适用于任意几何阵列^[15]。Xue等人为了解决IAA实时性较差的问题^[16],利用合适的Gohberg-Semencul型因子分解表示协方差矩阵并利用矩阵的Toeplitz结构来降低IAA的计算成本。Chen等人^[17]提出了一种Selective-IAA,该算法忽略了空域内的非感兴趣区域,只考虑感兴趣区域内的DOA估计,一定程度上提高了算法实时性。揭允康等人^[18]提出了一种高精度离网格DOA估计算法。该算法考虑到实际问题中目标位置与预定义离散网格的失配问题,对IAA功率谱进行修正,在角度粗估计后计算功率分量和偏移量,并进行交替优化,实现了高精度DOA估计。

在毫米波雷达目标探测的场景下,常遇到天线阵列孔径受限、信号相干、快拍数据少等问题。因此,本文针对这些问题提出了一种基于双平行线阵和联合迭代自适应算法与旋转不变技术(iterative adaptive approach and rotational invariance techniques, IAA-RIT)的高精度2D-DOA估计方法。在本文的内容安排中,首先在第1章中建立了TDM-MIMO模式下基于双平行线阵的雷达回波信号模型;其次,在第2章中基于第1章中的雷达数据立方,依次对回波信号进行Range-Doppler谱生成及多普勒相位补偿,并详细介绍了IAA-RIT的算法原理;在第3章中,设置了三组仿真实验来验证IAA-RIT的算法有效性;最后,在第4章中,总结了本文的主要工作内容。

1 信号模型

在天线阵元数量一定时,为了增加阵列孔径以获得更高的角分辨力,采用TDM-MIMO体制获得虚拟等效阵列。图1为TDM-MIMO体制下,LFMCW信号的波形示意图。

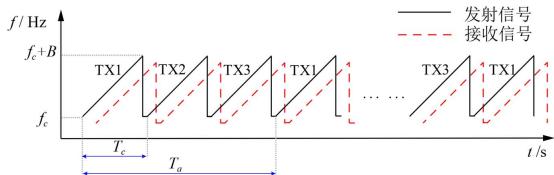


图1 TDM-MIMO体制下的LFMCW信号波形图

考虑到天线孔径受限且需要实现2D-DOA估计，本文选择一种双平行线阵列。相比较于均匀面阵，双平行线阵列减少了阵元个数。同时使用TDM技术，可以在阵元个数一定时，增大阵列孔径。图2为天线阵列示意图及其等效虚拟阵列。

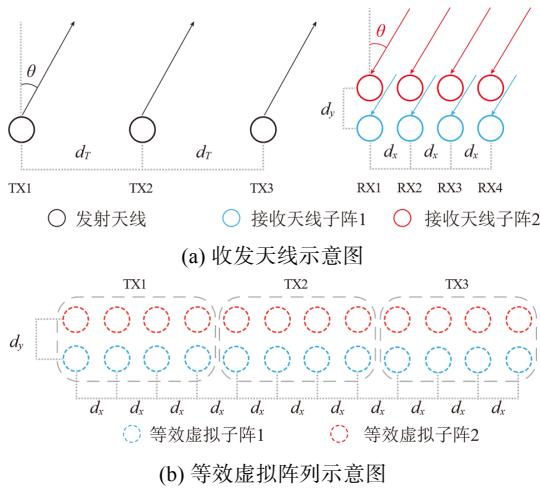


图2 收发天线及其虚拟等效阵列示意图

基于上述天线阵列，假设发射天线个数为 N_T ，接收天线个数为 N_R ，每个发射天线在每帧中共发射 N_C 个chirp信号，每个chirp中的采样点数为 N_S ，采样频率为 F_S 。假设目标匀速运动且速度较低，由目标运动引起的距离徙动小于一个距离单元。则子阵1的第 r 个接收天线接收到的来自第 t 个发射天线发射的第 n_c 个chirp的离散时间复指数信号可以由以下模型来表示：

$$x_{t,r}^1(n_c, n_s) = s_{t,r}(n_c, n_s) + e_{t,r}(n_c, n_s) \approx \\ \sum_{k=1}^K A_k e^{\phi_k} \exp \left[j2\pi \left(\frac{2R_k}{c} S + \frac{2v_k}{c} f_c \right) \frac{n_s - 1}{F_S} \right] \\ \exp[j2\pi(n_c - 1)T_a] \cdot \exp \left[j2\pi(t-1) \frac{2v_k}{c} f_c T_c \right] \\ \exp \left[j2\pi \left(\frac{r+tN_R/2}{\lambda} \right) d_x \cos \alpha_k \right] + e_{t,r}(n_c, n_s) \quad (1)$$

其中， $x_{t,r}^1(n_c, n_s)$ 表示子阵1的接收信号，

$t = 1, 2, \dots, N_T$, $r = 1, 2, \dots, N_R$, $n_c = 1, 2, \dots, N_C$, N_C 表示一帧内的 chirp 个数， $n_s = 1, 2, \dots, N_S$ ， N_S 表示每个 chirp 的采样点数， $e_{t,r}(n_c, n_s)$ 表示加性高斯白噪声，其均值为 0，方差为 σ^2 ， A_k 表示信号采样的复幅值， ϕ_k 表示信号固定相位， R_k 、 v_k 分别表示第 k 个目标的径向距离和速度， S 表示线性调频连续波的调频率。

子阵2的接收信号 $x_{t,r}^2(n_c, n_s)$ 与子阵1的接收信号仅在第5个指数项中略有不同， $x_{t,r}^2(n_c, n_s)$ 的第5个指数项为 $\exp \left[j2\pi \left(\frac{r+tN_R/2}{\lambda} \right) d_x \cos \alpha_k + d_y \cos \beta_k \right]$ ，第5个指数项中的 α 、 β 与目标的方位角 θ 、俯仰角 φ 之间的几何关系如图3所示。

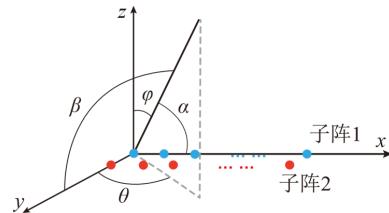


图3 α 、 β 与 θ 、 φ 的几何关系图

公式表示为：

$$\begin{aligned} \cos \alpha &= \sin \theta \sin \varphi \\ \cos \beta &= \cos \theta \sin \varphi \end{aligned} \quad (2)$$

2 高精度2D-DOA估计方法

2.1 生成Range-Doppler谱

由式(1)中的采样数据，可以获得雷达数据立方，其三个维度分别为ADC采样数、chirp个数、等效虚拟天线个数。对每个接收天线的ADC和chirp维度的原始回波数据进行2D-FFT，并沿着天线维进行非相参积累，对所得数据进行恒虚警检测（constant false alarm rate, CFAR）处理及峰值搜索，可获得有目标存在的Range-Doppler单元。沿天线维分别提取这些单元内的RD域数据，每个单元内的数据可表示为：

$$X_{t,r}^1 = \sum_{m=1}^M A_m e^{j\phi_m} P_{N_S} \left(\frac{2\pi n_{s,p}}{N_S} - \omega_R \right) P_{N_C} \left(\frac{2\pi n_{c,p}}{N_C} - \omega_v \right) \\ \exp \left[j2\pi(t-1) \frac{2v_k}{c} f_c T_c \right] \\ \exp \left[j2\pi \left(\frac{r+tN_R/2}{\lambda} \right) d_x \cos \alpha_k \right] + e_{t,r}(n_c, n_s) \quad (3)$$

式(3)中, $X_{t,r}^1$ 表示沿子阵 1 所提取的单个峰值单元处的数据。其中, M 表示在所提取的 RD 域峰值单元内存在 M 个目标; A_m 为第 m 个目标的信号采样复幅值; ϕ_m 为固定相位; $n_{s,p}$ 为距离维峰值位置索引, $0 \leq n_{s,p} < N_S$; $n_{c,p}$ 为多普勒维峰值位置索引, $0 \leq n_{c,p} < N_C$ 且有:

$$P_N(\omega) = \sum_{n=0}^{N-1} e^{-jn\omega} = \frac{\sin(\omega N/2)}{\sin(\omega/2)} e^{-j\omega(N-1)/2} \quad (4)$$

$$\omega_R = \frac{2\pi}{F_S} \left(\frac{2v_m}{c} f_c + \frac{2R_m}{c} S \right) \quad (5)$$

$$\omega_v = \frac{4\pi v_m T_c}{\lambda} \quad (6)$$

经过相同处理后, 子阵 2 的数据 $X_{t,r}^2$ 与 $X_{t,r}^1$ 仅在第 3 个指数项位置略有不同, $X_{t,r}^2$ 的第 3 个指数项为 $\exp\left[j2\pi\left(\frac{r+tN_R/2}{\lambda}\right)d_x \cos\alpha_k + d_y \cos\beta_k\right]$ 。

通过 2D-FFT 生成 RD 谱, 并经过 CFAR 处理, 不同距离不同速度的目标得到充分分离。后续的角度维信息处理仅在同距同速的 RD 单元内进行, 这不仅降低了 DOA 估计的难度, 同时避免了角度与距离、速度的额外配对。另外, 基于已知的目标速度, 对角度维信息进行相位补偿, 可提高 DOA 估计精度。

2.2 多普勒相位补偿

当 MIMO 发射天线工作在 TDM 模式时, 在一个 chirp 的持续时间内, 由于目标运动会产额外的相移。为了获得更好的 DOA 估计结果, 需要提前对角度维数据进行多普勒相位补偿。

将式(3)转换为矩阵形式, 可表示为:

$$X_0^1 = \mathbf{A}_0 \mathbf{s} + \mathbf{e}_0 \quad (7)$$

$$\mathbf{A}_0 = [\mathbf{a}_0(\alpha_1), \mathbf{a}_0(\alpha_2), \dots, \mathbf{a}_0(\alpha_m)] \quad (8)$$

$$\mathbf{a}_0(\alpha_m) = \mathbf{a}(\alpha_m) \oplus \mathbf{b} \quad (9)$$

$$\mathbf{a}(\alpha_m) = \left[1, e^{\frac{j2\pi}{\lambda}d_x \cos\alpha_m}, \dots, e^{\frac{j2\pi}{\lambda}(N_T N_R - 1)d_x \cos\alpha_m} \right]^T \quad (10)$$

$$\mathbf{b} = \left[[1, \dots, 1]_{N_R \times 1}, \left[e^{-j2\pi \frac{2v_m}{c} f_c T_c}, \dots, e^{-j2\pi \frac{2v_m}{c} f_c T_c} \right]_{N_R \times 1}, \dots, \left[e^{-j2\pi(N_T - 1)\frac{2v_m}{c} f_c T_c}, \dots, e^{-j2\pi(N_T - 1)\frac{2v_m}{c} f_c T_c} \right]_{N_R \times 1} \right]^T \quad (11)$$

$$\mathbf{s} = [s_1, s_2, \dots, s_m]^T \quad (12)$$

其中, 式(9)中 \oplus 表示 Hadamard 积; 式(11)为动目标引起的额外相移; 式(12)表示 m 个目标的信号复幅值。定义多普勒补偿向量为:

$$\tilde{\mathbf{b}} = \left[[1, \dots, 1]_{N_R \times 1}, \left[e^{-j2\pi \frac{2\tilde{v}_m}{c} f_c T_c}, \dots, e^{-j2\pi \frac{2\tilde{v}_m}{c} f_c T_c} \right]_{N_R \times 1}, \dots, \left[e^{-j2\pi(N_T - 1)\frac{2\tilde{v}_m}{c} f_c T_c}, \dots, e^{-j2\pi(N_T - 1)\frac{2\tilde{v}_m}{c} f_c T_c} \right]_{N_R \times 1} \right]^T \quad (13)$$

其中, \tilde{v}_m 表示第 m 个目标的速度估计值。对式(7)数据进行相位补偿, 表示为:

$$X^1 = X_0^1 \oplus \tilde{\mathbf{b}} = \mathbf{A}\mathbf{s} + \mathbf{e} \quad (14)$$

$$\mathbf{A} = [\mathbf{a}(\alpha_1), \mathbf{a}(\alpha_2), \dots, \mathbf{a}(\alpha_m)] \quad (15)$$

子阵 2 经过相同的处理后, 其数据表示为如下矩阵形式:

$$X^2 = \mathbf{A}\Phi\mathbf{s} + \mathbf{e} \quad (16)$$

$$\Phi = \text{diag}\left(e^{\frac{j2\pi}{\lambda}d_y \cos\beta_1}, e^{\frac{j2\pi}{\lambda}d_y \cos\beta_2}, \dots, e^{\frac{j2\pi}{\lambda}d_y \cos\beta_m}\right) \quad (17)$$

子阵 1 与子阵 2 均进行相位补偿后, 可进行后续的 DOA 估计。

2.3 联合 IAA 与 RIT 的高精度 2D-DOA 估计方法

为了能在阵列孔径受限、信号相干、少快拍的情况下进行高精度 2D-DOA 估计, 本文提出了一种基于双平行线阵列的联合 IAA-RIT 的 2D-DOA 估计方法。该方法可以有效克服相干信号的影响, 并且在少快拍时表现良好, 同时该算法无需进行额外的角度匹配, 减少了配对误差。联合 IAA-RIT 计算过程为: 首先, 基于双平行线阵列中的子阵 1, 使用 IAA 迭代更新信号协方差矩阵与能量矩阵, 待满足迭代条件后进行谱峰搜索, 计算出目标水平角、俯仰角的耦合角度估计值 $\tilde{\alpha}_m$; 接着, 利用子阵 2 与子阵 1 所得数据构造扩展的信号协方差矩阵 \mathbf{R} , 在处理相干信号时, 可利用空间平滑技术进行信号解相干; 最后, 利用耦合角度估计值 $\tilde{\alpha}_m$ 重构信号方向矩阵 $\tilde{\mathbf{A}}$, 利用 $\tilde{\mathbf{A}}$ 与 \mathbf{R} 计算子阵 1 与子阵 2 之间的旋转不变关系, 获得另外的耦合角度 $\tilde{\beta}_m$, 并利用式(2)进行角度解耦, 获得目标的水平角和俯仰角估计值。在详细介绍联合 IAA-RIT 之前,

将式(14)(16)中的信号模型重新定义为第 n 次快拍数据, 即

$$\mathbf{X}^1(n)=\mathbf{As}(n)+\mathbf{e} \quad (18)$$

$$\mathbf{X}^2(n)=\mathbf{A}\Phi\mathbf{s}(n)+\mathbf{e} \quad (19)$$

其中, $n=1,2,\cdots,N$, N 表示快拍总数。

下面将分三部分展开介绍联合 IAA-RIT 的具体内容。

第一部分, 基于子阵 1 中数据使用 IAA 进行 DOA 估计。

首先, 定义信号的能量对角矩阵 \mathbf{P} , 信号的协方差矩阵 \mathbf{R} , 信号之外的干扰和噪声的协方差矩阵 \mathbf{Q} , 分别表示如下:

$$P_m=\frac{1}{N} \sum_{n=1}^N |s_m(n)|^2 \quad (20)$$

$$\mathbf{R}=\frac{1}{N} \mathbf{X}^1 \mathbf{X}^{1H}=\mathbf{A}^1 \mathbf{P} \mathbf{A}^{1H} \quad (21)$$

$$\mathbf{Q}(\alpha_m)=\mathbf{R}-P_m \mathbf{a}(\alpha_m) \mathbf{a}^H(\alpha_m) \quad (22)$$

其中, P_m 为 \mathbf{P} 的第 m 个对角元素。定义加权最小二乘代价函数为:

$$\sum_{n=1}^N \| \mathbf{X}^1(n)-s_m(n) \mathbf{a}(\alpha_m) \|_W^2 \quad (23)$$

其中, $\|x\|_W^2 = \mathbf{x}^H \mathbf{W} \mathbf{x}$ 。当加权矩阵 $\mathbf{W}=\mathbf{Q}^{-1}(\theta_k)$ 时, 最小二乘估计量 $\tilde{s}_m(n)$ 的估计误差最小。由矩阵求逆引理可得, $\tilde{s}_m(n)$ 可表示为:

$$\tilde{s}_m(n)=\frac{\mathbf{a}^H(\alpha_m) \mathbf{R}^{-1} \mathbf{X}^1(n)}{\mathbf{a}^H(\alpha_m) \mathbf{R}^{-1} \mathbf{a}(\alpha_m)} \quad (24)$$

其次, 要计算出信号的空域功率谱, 需要对空域设置预定义离散网格字典, 设置网格数为 L ($L \gg M$)。接着对功率矩阵 \mathbf{P} 中的对角元素进行初始化, 可定义为:

$$\tilde{P}_l=\frac{1}{(\mathbf{a}^H(\alpha_l) \mathbf{a}(\alpha_l)) N} \sum_{n=1}^N |\mathbf{a}^H(\alpha_l) \mathbf{X}^1(n)|^2 \quad (25)$$

其中, $l=1,2,\cdots,L$ 。

最后, 使用式(20)(21)(24)对协方差矩阵 \mathbf{R} 及功率矩阵 \mathbf{P} 进行迭代更新, 直至达到收敛条件或最大迭代次数。本文中设置最大迭代次数为 15, 收敛条件为 $\left\|\operatorname{vec}\left(\mathbf{P}^{(t+1)}-\mathbf{P}^{(t)}\right)\right\|_2 / \left\|\operatorname{vec}\left(\mathbf{P}^{(t)}\right)\right\|_2 < \varepsilon$, $\varepsilon=1E-3$ 。迭代结束后, 通过谱峰搜索并利用索引值可计算耦合角度估计值 $\tilde{\alpha}_m$ 。

第二部分, 利用子阵 2 与子阵 1 所得数据构造扩展协方差矩阵 \mathbf{R} 。

首先, 定义子阵 1 的接收信号自相关矩阵为 \mathbf{R}_{11} , 子阵 2 与子阵 1 之间的互相关矩阵为 \mathbf{R}_{21} 。考虑到噪声自身的独立性, 且独立于信号, 则 \mathbf{R}_{11} 、 \mathbf{R}_{21} 分别表示为

$$\begin{aligned} \mathbf{R}_{11} &= E[\mathbf{X}^1 \mathbf{X}^{1H}]=\mathbf{A} \mathbf{P} \mathbf{A}^H+\sigma^2 \mathbf{I} \\ \mathbf{R}_{21} &= E[\mathbf{X}^2 \mathbf{X}^{1H}]=\mathbf{A} \boldsymbol{\Phi} \mathbf{P} \mathbf{A}^H \end{aligned} \quad (26)$$

其中, $\mathbf{P}=E[\mathbf{s} \mathbf{s}^H]$ 为信号协方差矩阵; σ^2 为加性高斯白噪声的方差; \mathbf{I} 为单位矩阵。

对 \mathbf{R}_{11} 进行特征分解, 令 $\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_M$ 为矩阵 \mathbf{R}_{11} 的 M 个较大特征值, 在白噪声的假设下, 可以由 $N_T N_R - M$ 个较小特征值的平均得到噪声方差的估计。

通过去除噪声的影响可得到:

$$\mathbf{C}_{11}=\mathbf{A} \mathbf{P} \mathbf{A}^H=\mathbf{R}_{11}-\sigma^2 \mathbf{I} \quad (27)$$

定义扩展协方差矩阵, $\mathbf{R} \in \mathbb{C}^{N_T N_R \times N_T N_R}$, 且

$$\mathbf{R}=\mathbf{R}_{21} \mathbf{C}_{11}^+ \quad (28)$$

其中, $\mathbf{C}_{11}^+=\mathbf{C}_{11}^H(\mathbf{C}_{11} \mathbf{C}_{11}^H)^{-1}$ 。

如果 \mathbf{A} 和 \mathbf{P} 为满秩矩阵且 $\boldsymbol{\Phi}$ 中对角元素均不相同, 那么 \mathbf{R} 的 M 个非零特征值等于 $\boldsymbol{\Phi}$ 中的 M 个对角元素, 而这些值对应的特征向量等于相应的信号方向向量, 即

$$\mathbf{RA}=\mathbf{A} \boldsymbol{\Phi} \quad (29)$$

若存在相干信号, 可使用文献[19]中的空间平滑技术重构矩阵 \mathbf{R}_{11} 、 \mathbf{R}_{21} 、 \mathbf{R} 。

然后, 基于 IAA 结果所获角度估计值 $\tilde{\alpha}_m$, 利用式(10)与式(15)重构信号方向矩阵 $\tilde{\mathbf{A}}$,

$$\tilde{\mathbf{A}}=[\mathbf{a}(\tilde{\alpha}_1), \mathbf{a}(\tilde{\alpha}_2), \cdots, \mathbf{a}(\tilde{\alpha}_m)] \quad (30)$$

$$\mathbf{a}(\tilde{\alpha}_m)=\left[1, e^{j \frac{2 \pi}{\lambda} d_x \cos \tilde{\alpha}_m}, \cdots, e^{j \frac{2 \pi}{\lambda}(N_T N_R-1) d_x \cos \tilde{\alpha}_m}\right]^T \quad (31)$$

最后, 计算旋转不变因子 $\boldsymbol{\Phi}$ 。

$$\boldsymbol{\Phi}=\tilde{\mathbf{A}}^H \mathbf{R} \tilde{\mathbf{A}} \quad (32)$$

其中, $\tilde{\mathbf{A}}^+=\tilde{\mathbf{A}}^H(\tilde{\mathbf{A}} \tilde{\mathbf{A}}^H)^{-1}$ 。利用式(17)可获得耦合角度估计值 $\tilde{\beta}_m$ 。

第三部分, 利用耦合角度估计值 $\tilde{\alpha}_m$ 、 $\tilde{\beta}_m$ 以及它们与水平角和俯仰角之间的空间几何关系进行角度解耦。

利用式(2)求解方位角 θ_m 与俯仰角 φ_m , 表

达式如下：

$$\begin{aligned}\theta_m &= \arctan(\cos \tilde{\alpha}_m / \cos \tilde{\beta}_m) \\ \varphi_m &= \arcsin(\sqrt{\cos^2 \tilde{\alpha}_m + \cos \tilde{\beta}_m})\end{aligned}\quad (33)$$

至此，已完成高精度 2D-DOA 估计。

3 仿真实验

为了验证本文所提出的基于 IAA-RIT 的 DOA 估计性能，本文设置了三组不同的仿真实验分别观察 IAA-RIT 在不同信噪比不同快拍数下的分辨力、精度及分辨相干信号的能力。实验假设雷达静止不动，目标以固定速度和角度低速运动且等效虚拟阵列为双平行线阵列。

3.1 IAA-RIT 角度分辨率性能验证实验

为了验证本文所提出的 IAA-RIT 在少快拍情况下的角度分辨能力，将分别使用 IAA-RIT、2D-MUSIC、2D-ESPRIT 算法进行仿真实验，并对实验结果进行对比。实验中设置快拍数 $N=10$ ，设置两组不同信噪比为 25 dB、15 dB，设置目标角度为 $(-20^\circ, 5^\circ), (-5^\circ, -10^\circ)$ 。实验结果如图 4 所示。

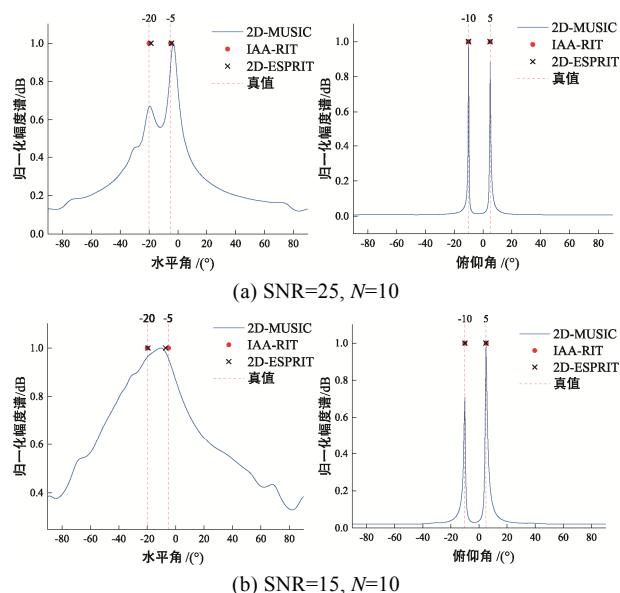


图 4 不同算法的 2D-DOA 估计结果图

由图 4(a)中实验结果可知，在信噪比为 25 dB 时，尽管快拍数较少，三种算法均可分辨出两个目标。由图 4(b)可知，当信噪比下降为 10 dB 时，2D-MUSIC 算法在水平角度方向无法正确分辨目标，2D-ESPRIT 算法也在估计水平角时出现了较为明显的偏移，而本文所提出的 IAA-RIT 在水平

角和俯仰角维度上均表现出良好的估计性能，不仅正确分辨出目标，同时所计算的角度未出现明显的偏移。

图 4 中实验结果表明，本文所提出的 IAA-RIT 具有在极少快拍、信噪比较低时的目标分辨能力且角度估计精度高于 2D-MUSIC 算法和 2D-ESPRIT 算法。

3.2 IAA-RIT 高精度估计性能验证实验

为了验证本文所提出的 IAA-RIT 在少快拍时的 DOA 估计精度，使用 IAA-RIT 在不同快拍数和不同信噪下进行 200 次蒙特卡洛实验，并使用 2D-MUSIC 算法进行相同条件下的仿真实验作为对比。设置目标角度为 $(-20^\circ, 5^\circ), (-5^\circ, -10^\circ)$ 。实验运行结果使用均方根误差（root mean square error, RMSE）作为单个角度计算精度的评价指标，其计算公式为式(34)。实验结果如图 5 所示。

$$RMSE = \sqrt{\frac{1}{2KN_m} \sum_{k=1}^K \sum_{i=1}^{N_m} (\tilde{\theta}_{i,k} - \theta_k)^2 + (\tilde{\varphi}_{i,k} - \varphi_k)^2} \quad (34)$$

其中， K 为目标个数； N_m 为蒙特卡洛实验次数； $\tilde{\theta}_{i,k}$ 、 $\tilde{\varphi}_{i,k}$ 分别为进行第 i 次蒙特卡洛实验时的第 k 个目标的水平角估计值和俯仰角估计值； θ_k 、 φ_k 为第 k 个目标的水平角真实值和俯仰角真实值。

由图 5 中实验结果可知，使用 IAA-RIT 进行 2D-DOA 估计时，在快拍数低于 100 时，其计算结果的均方根误差均低于 2D-MUSIC 算法。在信噪比为 20 dB 时，尽管快拍数为 5、10，使用 IAA-RIT 所得角度的均方根误差仅分别为 0.190 98、0.306 15。由此可知，当信噪比良好时，IAA-RIT 在极少快拍数的情况下也能获得不错的估计精度。当信噪比为 5 dB

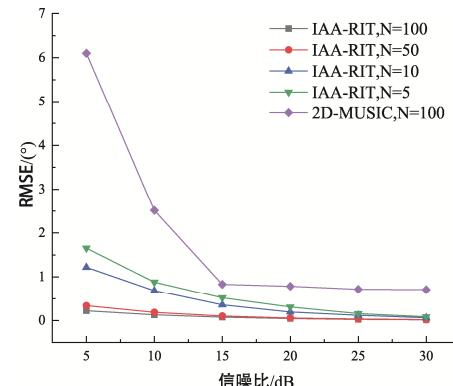


图 5 不同信噪比时 IAA-RIT 和 2D-MUSIC 的 RMSE

时, IAA-RIT 在快拍数为 50、100 时, 其所计算的角度均方根误差分别为 0.336 8、0.218 4, 由此说明当快拍数较多时, IAA-RIT 在信噪比较低的情况下可以获得不错的估计精度。

图 5 中实验结果表明, 本文所提出的 IAA-RIT 在低信噪比或低快拍数的情况下具有良好的 2D-DOA 估计精度。

3.3 IAA-RIT 相干信源角度分辨性能与估计精度验证实验

为了验证本文所提出的 IAA-RIT 在少快拍且存在信源相干情况下时的 2D-DOA 估计性能, 在本次仿真实验中, 将使用 IAA-RIT 进行 200 次蒙特卡洛实验, 观察算法的角度估计情况。实验设置目标角度为 $(-20^\circ, 15^\circ)$, $(5^\circ, -10^\circ)$ 。实验结果如图 6 所示。

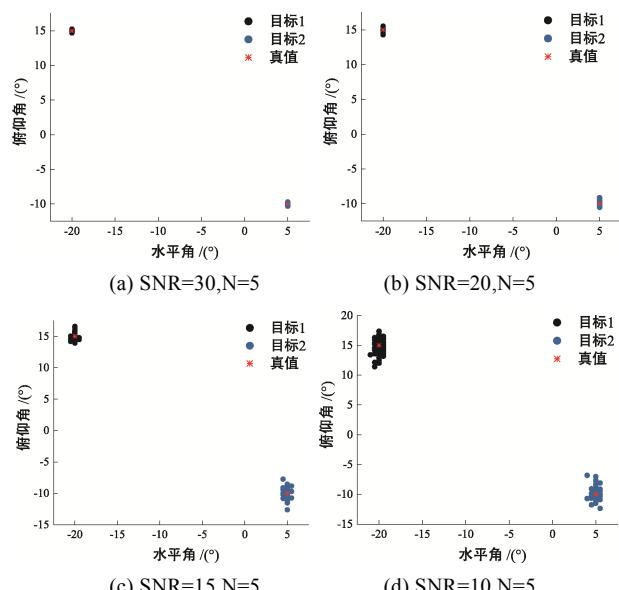


图 6 少快拍且信源相干时, 不同信噪比下 IAA-RIT 的 2D-DOA 估计结果

由图 6 中实验结果可知, IAA-RIT 可以在极少快拍下估计出相干信号的角度值。随着信噪比的不断提升, 蒙特卡洛实验结果的目标落点更加紧凑, 这说明在极少快拍下使用 IAA-RIT 进行 2D-DOA 估计的精度不断提高。实验结果表明, 当存在相干信号且快拍数较低时, IAA-RIT 可在信噪比良好时获得较高的 2D-DOA 估计性能。

通过以上三组实验可知, 基于双平行线阵列的 IAA-RIT 具有在快拍数据较少时的角度分辨能力, 且可以仅使用较少快拍数据在信噪比良好时对相

干信号进行高精度 2D-DOA 估计。

4 结束语

针对汽车毫米波雷达目标探测中可能出现的阵列孔径受限、信号相干、少快拍的情况, 本文提出了一种无需额外角度匹配的基于双平行线阵的联合迭代自适应与旋转不变技术(IAA-RIT)的 2D-DOA 估计方法。IAA-RIT 基于双平行线阵, 利用双平行线阵中的子阵 1 构建 IAA 谱, 并通过两个子阵构建扩展协方差矩阵, 并重构信号方向矩阵, 从而计算两个子阵间的旋转不变关系, 最后通过角度解耦估计出目标的水平角和俯仰角。实验结果表明, 该算法的 2D-DOA 估计性能高于相同阵列下的 2D-MUSIC 和 2D-ESPRIT。在信噪比良好时, IAA-RIT 仅需要极少快拍即可进行高精度 2D-DOA 估计, 并且存在相干信号时也能保证算法的有效性且无需进行额外的角度匹配。

参考文献:

- [1] JIANG M J, XV G, PEI H, et al. 4D high-resolution imagery of point clouds for automotive mmwave radar[J]. IEEE Transactions on Intelligent Transportation Systems, 2024, 25(1): 998-1012.
- [2] LI X R, WANG X D, YANG Q, et al. Signal processing for TDM MIMO FMCW millimeter-wave radar sensors[J]. IEEE Access, 2021, 9: 167959-167971.
- [3] ENGELS F, HEIDENREICH P, WINTERMANTEL M, et al. Automotive radar signal processing: research directions and practical challenges[J]. IEEE Journal of Selected Topics in Signal Processing, 2021, 15(4): 865- 878.
- [4] PATOLE S M, TORLAK M, WANG D, et al. Automotive radars: a review of signal processing techniques[J]. IEEE Signal Processing Magazine, 2017, 34 (2): 22-35.
- [5] SUN S Q, PETROPULU A P, POOR H V. MIMO radar for advanced driver-assistance systems and autonomous driving: advantages and challenges[J]. IEEE Signal Processing Magazine, 2020, 37(4): 98-117.
- [6] RAMBACH K, YANG B. MIMO radar: time division multiplexing vs. code division multiplexing[C]//International Conference on Radar Systems (Radar 2017), Belfast, 2017: 1-5.
- [7] DULY A J, LOVE D J, KROGMEIER J V. Time-division beamforming for MIMO radar waveform design[J]. IEEE Transactions on Aerospace and Electronic Systems, 2013, 49(2): 1210-1223.
- [8] 张泽宇. 毫米波 MIMO 雷达阵列优化设计与 DOA 估计方法研究[D]. 成都: 电子科技大学, 2023.

- [9] 刘献巨. 基于 MIMO 阵的车载毫米波雷达 DOA 估计技术研究[D]. 成都: 电子科技大学, 2023.
- [10] 单泽彪, 王宇祥, 常立民, 等. 冲击噪声背景下相干信号 DOA 估计[J]. 电子测量技术, 2022, 45(15): 166-171.
- [11] 周英钢, 邵佳伟. 对数螺旋阵列的相干信号 DOA 估计研究[J]. 电子测量与仪器学报, 2023, 37(2): 220-227.
- [12] MALIOUTOV D, CETIN M, WILLSKY A S. A sparse signal reconstruction perspective for source localization with sensor arrays[J]. IEEE Transactions on Signal Processing, 2005, 53(8): 3010-3022.
- [13] 李子高, 李淑秋, 闻疏琳. 一种改进的正交匹配追踪的 DOA 估计方法[J]. 测控技术, 2017, 36(1): 27-31.
- [14] 张宇, 景鑫磊, 蒋忠进. 一种基于稀疏贝叶斯学习的离网 DOA 估计算法[J]. 雷达科学与技术, 2024, 22(1): 35-42.
- [15] YARDIBI T, JIAN L, STOICA P, et al. Source locali-
- zation and sensing: a nonparametric iterative adaptive approach based on weighted least squares[J]. IEEE Transactions on Aerospace and Electronic Systems, 2010, 46(1): 425-443.
- [16] XUE M, XU L, LI J. IAA spectral estimation: fast implementation using the gohberg-semencul factorization[J]. IEEE Transactions on Signal Processing, 2011, 59(7): 3251-3261.
- [17] CHEN Y, HUANG L T, SO H C. Selective range iterative adaptive approach for high-resolution DOA estimation[J]. IEEE Access, 2019, 7: 15634-15640.
- [18] 揭允康, 张雯, 李想, 等. 一种基于迭代自适应的离网格 DOA 估计方法[J]. 电子与信息学报, 2023, 45(10): 3805-3811.
- [19] 郑超, 杨志强. 空间平滑 MUSIC 算法的相干信源 DOA[J]. 火控雷达技术, 2021, 50(3): 21-24.

(上接第 186 页)

参考文献:

- [1] 朱小波, 谭兴文. 自动驾驶汽车环境感知与传感器融合技术[J]. 汽车与新动力, 2024, 7(4): 24-27.
- [2] FAN L, WANG J, CHANG Y, et al. 4D mmwave radar for autonomous driving perception: a comprehensive survey[J]. IEEE Transactions on Intelligent Vehicles, 2024, 9(4): 4606-4620.
- [3] 曹璐. FMCW 毫米波雷达超分辨率测距算法研究[D]. 武汉: 华中科技大学, 2022.
- [4] 刘劭扬. FMCW 雷达传感器高精度测距算法研究及实现[D]. 成都: 电子科技大学, 2023.
- [5] 戴作宁. 毫米波雷达高速运动目标精确测速研究[D]. 南京: 南京大学, 2016.
- [6] 聂辉勇. 高速运动微弱目标参数估计算法研究[D]. 厦门: 厦门大学, 2021.
- [7] 黄岩, 张慧, 兰吕鸿康, 等. 汽车毫米波雷达信号处理技术综述[J]. 雷达学报, 2023, 12(5): 923-970.
- [8] 张宇乐, 周豪, 胡国平, 等. 稀疏阵列结构设计及波达方向估计研究进展[J]. 信号处理[2024-08-27].
- [9] 李荣禄, 汤建龙, 袁永强. 基于 TOEPLITZ 重构的压缩感知嵌套阵列 DOA 估计[J]. 雷达科学与技术, 2024, 22(3): 334-340.
- [10] CAI, J F, EMMANUEL J C, SHEN Z W. A singular value thresholding algorithm for matrix completion[J]. SIAM Journal on Optimization, 2010, 20(4): 1956-1982.
- [11] LIU J, WANG X P, ZHOU W D. Covariance vector sparsity-aware DOA estimation for monostatic MIMO radar with unknown mutual coupling[J]. Signal Processing, 2016, 119: 21-27.
- [12] LI P, RUI Y B, YUAN Y F, et al. A sparse array DOA es-
- timation approach based on matrix completion[C]// 2022 14th International Conference on Signal Processing Systems (ICSPS). IEEE, 2022.
- [13] ZHENG Z, HUANG Y X, WANG W Q, et al. Direction-of-arrival estimation of coherent signals via coprime array interpolation[J]. IEEE Signal Processing Letters, 2020, 27: 585-589.
- [14] FRANCESCO B, WIM V R, PETER H. 2D-MUSIC technique applied to a coherent FMCW MIMO radar[C]// IET International Conference on Radar Systems (Radar 2012), Glasgow, UK, 2012: 1-6.
- [15] 黄耀华. 二维 DOA 估计算法研究及实现[D]. 西安: 西安电子科技大学, 2022.
- [16] CAO H, LIU Q. Accurate DOA estimation based on real-valued singular value decomposition[C]// 2019 IEEE International Conference on Signal, Information and Data Processing (ICSIDP), Chongqing, China, 2019: 1-4.
- [17] ZHENG Z, HUANG Y X, WANG W Q, et al. Augmented covariance matrix reconstruction for DOA estimation using difference co-array[J]. IEEE Transactions on Signal Processing, 2021, 69: 5345-5358.
- [18] SUN B, WU C X, RUAN H L, et al. Direction-of-arrival estimation for coprime array via covariance matrix reconstruction[C]// 2020 IEEE 3rd International Conference on Electronics Technology (ICET). IEEE, 2020.
- [19] ZHANG X, ZHENG Z, WANG W Q, et al. DOA estimation of coherent sources using coprime array via atomic norm minimization[J]. IEEE Signal Processing Letters, 2022, 29: 1312-1316.
- [20] CHEN Z, FAN C Y, HUANG X T. Interpolation-based direction-of-arrival estimation for coprime arrays via covariance matrix fitting[J]. IEEE Access, 2020, 8: 149133-149141.

基于均方误差网络优化的时序 InSAR 相位解缠方法

王彦平¹, 王纪清¹, 白泽朝¹, 余快², 史海涛², 申文杰¹

1. 北方工业大学 信息学院, 北京 100144

2. 中国空间技术研究院, 北京 100094

摘要: 相位解缠是时序 InSAR (MT-InSAR) 高精度形变测量的关键步骤, 合适的空间网络结构直接影响相位解缠的精度和可靠性。目前主流的 Delaunay 三角网在大梯度形变的情况下易包含不满足相位连续性假设的弧段, 造成解缠误差。提出了一种基于均方误差 (MSE) 的空间网络优化方法。采用最小二乘法 (LS) 计算各弧段的解缠相位, 利用均方误差筛选弧段, 剔除较大均方误差值的弧段; 采用 Dijkstra 算法, 以时间相干性为权重, 重新连接被剔除弧段的顶点, 构建更加可靠的网络; 利用 Edgelist 算法空间维相位解缠。采用仿真数据对不同网络解缠结果进行了对比分析。实验结果表明, 相比于 Delaunay 网络, 解缠精度提升了 64.4%, 与 APSP 网络相比解缠精度提升了 6.5%, 验证了所提方法的有效性。

关键词: 相位解缠; 时序 InSAR; 空间网络优化; 最小二乘法; Dijkstra 算法

文献标志码: A

中图分类号: TP391

Phase Unwrapping Method in MT-InSAR Based on MSE Network Optimization

WANG Yanping¹, WANG Jiqing¹, BAI Zechao¹, YU Kuai², SHI Haitao², SHEN Wenjie¹

1. School of Information Science and Technology, North China University of Technology, Beijing 100144, China

2. China Academy of Space Technology, Beijing 100094, China

Abstract: Phase unwrapping is a crucial step in high-precision deformation measurement using multitemporal interferometric synthetic aperture radar (MT-InSAR). An appropriate spatial network structure directly affects the accuracy and reliability of phase unwrapping. The currently prevalent Delaunay triangulation tends to include arcs that do not satisfy the phase continuity assumption under high phase gradient conditions, resulting in unwrapping errors. This paper proposes a spatial network optimization method based on mean square error (MSE). First, the least squares (LS) method is employed to compute the unwrapped phase for each arc, and an initial arc selection is performed based on MSE to remove arcs with larger MSE values. Then, the Dijkstra algorithm is used to reconnect the removed arcs' vertices, using temporal coherence as the weighting factor, to construct a more reliable network. Finally, the Edgelist algorithm is applied for spatial phase unwrapping. A comparison analysis of unwrapping results from different networks is conducted using simulated data. The experimental results show that compared with the Delaunay network, the unwrapping accuracy is improved by 64.4%, and by 6.5% compared with the all-pairs-shortest-path (APSP) network, demonstrating the effectiveness of the proposed method.

Key words: phase unwrapping; MT-InSAR; spatial network optimization; least squares method; Dijkstra

时序 InSAR 技术凭借全天时全天候、高精度等的优势, 广泛应用于矿山、滑坡、桥梁等区域的

大范围、长时间形变监测领域^[1-4]。然而, 当发生大梯度形变时, InSAR 解算精度会降低^[4-5]。其中,

基金项目: 国家重点研发青年科学家项目 (2023YFB3905200); 北京市教育委员会科学研究计划项目 (KM202410009001)。

作者简介: 王彦平 (1976—), 男, 博士, 研究员, 博士生导师, 研究方向为雷达成像安全监测技术、遥感智能信息处理及应用研究; 王纪清 (2000—), 男, 硕士研究生, 研究方向为 InSAR 相位解缠算法研究; 白泽朝 (1991—), 男, 博士, 讲师, 研究方向为星载/地基雷达形变测量方法与应用研究。

相位解缠是影响形变解算精度的关键步骤。InSAR 只能获取缠绕在 $(-\pi, \pi]$ 之间的相位，需要通过解缠还原真实相位^[6]。一般相位解缠要满足相位连续性假设，即相邻解缠点的相位梯度应小于 π 以保持相位的空间连续性^[7]。以某一像素点为解缠起点，其余点的真实相位可通过对缠绕相位梯度积分求得^[8-9]。但在大梯度形变区域，传统的相位连续性假设往往失效，导致解缠误差。

在时序 InSAR 相位解缠中，需要对识别的相干点进行稀疏构网，通常采用 Delaunay 三角网^[10]，以最大化最小角度的原则提高相位解缠的稳定性，但网络中仍会存在相邻解缠点的相位梯度不满足相位连续性假设的情况。近年来，学者们通过相干点目标的网络优化来提高相位解缠的精度^[11]。

Zhang 等人基于 LS 的模糊度检测器对弧上的相位差进行建模，并在出现异常值时将其删除^[12]。Song 等人提出一种自适应时序 InSAR 技术应用在桥梁监测上，利用均方误差识别桥梁伸缩缝，将网络划分为多个子网同时解缠，进一步提高解缠精度，但在每个子网中都需要选择参考点进行约束估计，增加计算负担^[13]。Luo 等人提出对筛选后网络中各个分级网络和孤立点之间进行重新连接形成的增强型 Delaunay 三角网络，获得一种不规则的网络，尽管连通了所有相干点，但降低了网络的整体连通性，影响了后续解缠的可靠性^[14]。马张烽等人提出利用弧段的时间相干性作为衡量时序干涉图中相位梯度平均水平的质量指标，构建 APSP (All-Pairs-Shortest-Path) 网络，优化网络结构，但实际应用中仅以时间相干性无法全面评估弧段质量^[15]。

对于 Delaunay 三角网，相邻点之间可能存在相位跃变，出现显著的相位不连续，而非相邻点间却有可能满足相位连续性假设。因此，本文提出了一种均方误差筛选的网络优化相位解缠策略，在现有时序 InSAR 方法基础上，引入均方误差优化相位解缠空间网络，提高解缠的精度。本文的结构安排为第 1 章介绍所提出空间网络优化相位解缠的方法和处理步骤；第 2 章采用仿真数据对本文方法与 Delaunay 网络和 APSP 网络的解缠结果进行分析对比；第 3 章对全文进行了总结。

1 空间网络优化的相位解缠方法

时序 InSAR 数据处理的主要步骤包括 SAR 影像的预处理、相干点选取、网络构建、相位解缠以及形变反演，如图 1 所示。本文针对时序 InSAR 处理流程中的构网以及相位解缠环节进行优化。首先，引入均方误差识别 Delaunay 三角网的弧段相位跃变位置，筛选出不可靠的弧段；其次引入 Dijkstra 最短路径算法，以时间相干性作为孤立点连接主网络的距离权重，形成可靠的相干点网络；最后，使用 Edgelist 算法进行相位解缠^[16]。

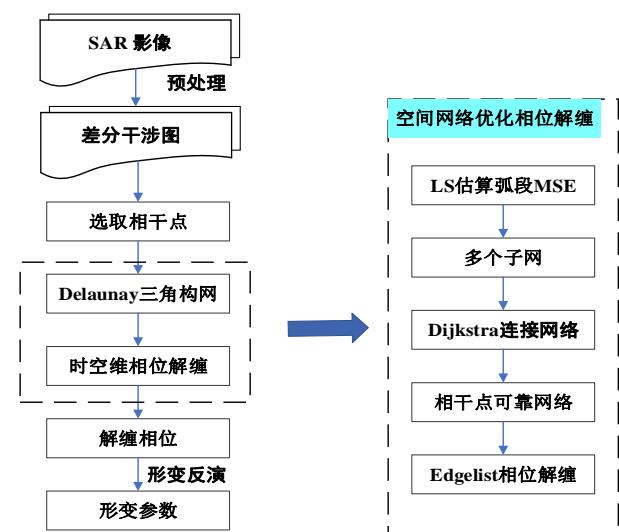


图 1 时序 InSAR 相位解缠优化处理流程

1.1 基于均方误差筛选的可靠网络构建

在大形变与噪声共同影响下，差分干涉图中可能出现低相干性区域，这些区域的相位信息通常会出现不连续性。由于 Delaunay 三角网络的构建仅基于点的空间位置，未能充分考虑相位不连续区域的影响。因此，需要对 Delaunay 三角网络进行改进。首先，以弧上两点的相位差作为观测值，构建相位观测模型，引入均方误差筛选发生相位跃变的弧，降低相位梯度过大的弧段占所有弧段的比例。

假设有 N 幅 SAR 图像，生成 M 幅差分干涉图，在第 i 幅差分干涉图上，弧段的相位差可以表示为：

$$\Delta\psi_{x,y}^i = \Delta\varphi_{x,y}^i + 2k_{x,y}^i\pi \quad (1)$$

x 和 y 分别代表弧段两端点， $\Delta\psi_{x,y}^i$ 为弧段在第 i 幅差分干涉图的真实相位梯度， $\Delta\varphi_{x,y}^i$ 为弧段在第 i 幅差分干涉图的缠绕相位梯度， $k_{x,y}^i$ 为弧段

的相位模糊度。假设弧段两端之间的真实相位差的绝对值不超过 π ，则可以对缠绕相位重新缠绕来恢复出真实相位。式(1)可以重写为：

$$\text{wrap}(\Delta\phi_{x,y}^i) = \Delta\psi_{x,y}^i \quad (2)$$

$\text{wrap}(\cdot)$ 表示为缠绕操作。 N 幅 SAR 图像是在有序的时间序列 $(t_0, t_1, \dots, t_{N-1})$ 生成 M 幅差分干涉图，弧段上的相位双差可以用弧段相位时间序列矢量

$\Delta\Phi_{x,y} = [\Delta\phi_{x,y}^{(t_0,t_1)}, \Delta\phi_{x,y}^{(t_1,t_2)}, \dots, \Delta\phi_{x,y}^{(t_{N-2},t_{N-1})}]$ 表示，假设形成第 i 幅差分干涉图的两个图像的获取日期为 t_s 和 t_e ($t_s < t_e$)，该弧段的相位双差表示为：

$$\Delta\psi_{x,y}^{(t_s,t_e)} = \Delta\phi_{x,y}^{(t_s,t_{s+1})} + \Delta\phi_{x,y}^{(t_{s+1},t_{s+2})} + \dots + \Delta\phi_{x,y}^{(t_{e-1},t_e)} \quad (3)$$

进一步可以得到：

$$\text{wrap}(\Delta\phi_{x,y}) = A\Delta\Phi_{x,y} + \mathbf{n} \quad (4)$$

$$\mathbf{y} = Ax + \mathbf{n} \quad (5)$$

其中， A 是稀疏的系数矩阵，大小为 $M \times N-1$ ， \mathbf{n} 为相位不闭合的去相关噪声。

使用最小二乘法估算出 $\hat{\Delta\Phi}_{x,y}$ 。

$$\hat{\Delta\Phi}_{x,y} = \hat{x} = (A^T PA)^{-1} A^T Py \quad (6)$$

$$V = Ax - y \quad (7)$$

其中， V 的大小为 $M \times 1$ ，对应该弧段在每一幅差分干涉图中的残差， P 为观测值的权重矩阵，若弧段两端之间的真实相位差的绝对值超过 π ，则 $\text{wrap}(\Delta\phi_{x,y}^{(t_s,t_e)}) \neq \Delta\psi_{x,y}^{(t_s,t_e)}$ ，代表着弧段具有相位跃变，这里采用均方误差作为相位跃变的粗筛指标，表示为：

$$MSE = \sqrt{\frac{V^T PV}{M-N-1}} \quad (8)$$

此时，粗筛后整个相干点网络被划分为多个子网络。为了确保网络连通，需要对网络进行二次连接。弧段质量的另一个度量标准可以用时间相干性表示为：

$$\gamma_{x,y} = \frac{\left| \sum_{i=1}^M e^{j\Delta\phi_{x,y,i}} \right|}{M} \quad (9)$$

其中， M 为差分干涉图的数量，可以看出弧段 (x,y) 的时间相干性 $\gamma_{x,y}$ 是由其缠绕相位梯度求复数模叠加取平均计算得到，反映了弧段相位梯度的平均水平，时间相干性越大代表相位梯度越小。为了包含尽可能多的可解缠的相干点，且保持整体网络的

连通性，引入 Dijkstra 最短路径算法，通过时间相干性选择两点之间的最短路径替代已删除的弧段。

Dijkstra 最短路径算法是一种用于计算加权图中单源最短路径的经典算法。设无向图 $G=(V,E)$ ，其中 V 是节点集合， E 是边的集合。设源点为 $s \in V$ 。为了计算从源点 s 到图中其他所有顶点的最短路径，定义一个距离函数：

$$dist(s) = 0, dist(v) = \infty, v \in V \text{ 且 } v \neq s \quad (10)$$

设置源点到自身的距离为 0，到其他所有点的初始距离为无穷大，将所有顶点标记为“未访问”，使用一个优先队列存储访问节点，并按照节点的当前最短距离排序。

对于每一个从当前节点 u 指向其邻接节点 v 的边 (u,v) ，在路径搜索中以权重 $w(u,v)$ 保持松弛操作更新最短路径，松弛操作表达为：

$$dist(v) = \min(dist(v), dist(u) + w(u,v)) \quad (11)$$

松弛操作结束后，获得起始节点到邻接节点由多条边组成的最短路径。重新连接网络时，将所有相干点作为顶点集合 V ，筛除的弧段作为边集合 E 。弧段的平均相位梯度水平越低，时间相干性越高，距离权重则越小。为了减少算法复杂度，采用 KNN 算法去识别顶点邻近的 K 个点构成弧段作为集合 E^+ 。随着搜索进行，筛除的边 E 会被 E^+ 里面的较低距离权重的边所替代连接到主网络。此时更新完所有弧段，新生成的网络相对于 Delaunay 网络消除了大部分不满足相位连续性假设的弧段，同时保证了网络的连通性，总的网络质量更高。

1.2 Edgelist 相位解缠

改进后的网络破坏了原有的三角形拓扑结构，导致传统依赖独立三角形约束的最小费用流 (minimum cost flow, MCF) 求解器难以有效地实现相位解缠。为解决这一问题，采用了 Edgelist 相位解缠算法^[16]，将约束从三角形转换为弧段。这种转换方法使得可以重新构建 MCF 目标函数，从而成功地获取每幅差分干涉图的解缠相位。

由(1)(2)式可知，相干点缠绕相位和相位模糊度的关系可以表示如下：

$$\frac{\Delta\phi_{x,y} - \text{wrap}(\Delta\phi_{x,y})}{2\pi} = n_y - n_x + k_{x,y} \quad (12)$$

其中， n_y 和 n_x 是相干点的缠绕相位和解缠相位相差的整数模糊度， $k_{x,y}$ 为弧段 (x,y) 的相位模糊度。

由于 $|\text{wrap}(\Delta\varphi_{x,y})| < \pi$, 式(12)可以改写为:

$$n_y - n_x + k_{x,y} = \left[\frac{\Delta\varphi_{x,y}}{2\pi} \right] = \left[\frac{\varphi_x - \varphi_y}{2\pi} \right] \quad (13)$$

其中, $[\cdot]$ 代表取整操作, 式(13)作为约束条件, 相位解缠问题转化为求解约束条件下目标函数最小化的问题。目标函数表示为:

$$\min \left\{ \sum_{(x,y)}^{N-1} w_{x,y} |k_{x,y}| \right\} \quad (14)$$

其中, $w_{x,y}$ 表示权重因子, 目标函数可以用线性规划进行求解。

2 仿真实验与分析

2.1 仿真数据

为了验证所提出的网络优化方法的有效性, 本文采用了空间相干性定权的 EdgeList 相位解缠算法, 利用仿真数据对 Delaunay 网络、仅以时间相干性衡量网络质量的 APSP 网络以及所提出的优化网络进行空间维解缠对比分析。本文构建了 2000×2000 像素的仿真区域, 区域内随机生成了 10 000 个相干点。同时生成了 24 幅差分干涉图, 图 2(a)为利用 peaks 函数模拟地表运动的形变速率图。为了进一步增加模拟数据的复杂性和真实性, 利用分型模型为每一幅差分干涉图模拟大气相位, 并加入了标准差为 0.8 的高斯噪声, 以模拟实际观测中的相位噪声。在仿真数据中, 第十幅差分

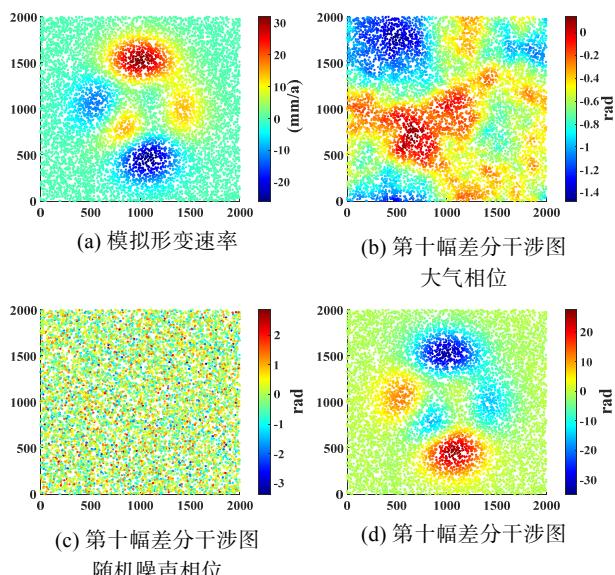


图 2 仿真数据

干涉图时间基线最长, 相位变化最大。如图 2(b)为第十幅差分干涉图的大气相位, 图 2(c)为第十幅差分干涉图的随机相位噪声, 图 2(d)表示包含了形变、大气、相位噪声的第十幅差分干涉图模拟相位。

2.2 相位解缠结果与分析

图 3 展示了不同网络在相位解缠性能方面的对比。图 3(a)、(d) 和 (g) 为第十幅差分干涉图的解缠结果。为了直观评估解缠效果, 将所得的解缠相位与模拟的真实相位进行差分, 得到相位误差分布图, 如图 3(b)、(e) 和 (h) 所示, 清晰地显示了解缠相位与真实相位之间的差异。除此之外, 图 3 还比较了不同网络在时间序列上的累计解缠误差, 计算了时序解缠相位图(共 24 景)的模糊度误差总和, 其中每个模糊整数对应 2π 的解缠误差, 结果如图 3(c)、(f) 和 (i) 所示。

对图 3 结果进行分析, 传统的 Delaunay 三角形网络在形变程度不同的区域, 其相位解缠误差存在明显差异, 误差主要集中在形变较大的区域。相比之下, APSP 网络和所提出的网络整体解缠效果一致, 有效地减轻了大形变区域的解缠误差, 使得误差分布更加稀疏。此外, 对比 APSP 网络的累计解缠误差时发现, 其在个别点的误差总和超过了所提出的网络。

表 1 不同网络解缠结果 RMSE 对比

Delaunay	APSP	提出网络
RMSE/rad	0.607	0.231

为了进一步定量分析三种网络解缠结果的精度, 这里用解缠相位的 RMSE 来进行衡量。解缠相位的 RMSE 定义为:

$$RMSE = \sqrt{\frac{\sum_{i=1}^P (\hat{\psi}_i - \psi_i)^2}{P}} \quad (15)$$

其中, $\hat{\psi}_i$ 和 ψ_i 分别表示为相干点的解缠相位和真实相位, P 为相干点的数量。

如表 1 所示, 分别得到三个网络的解缠精度为 0.607 rad、0.231 rad 和 0.216 rad。所提出的网络相对于 Delaunay 和 APSP 网络精度分别提升 64.4% 和 6.5%。

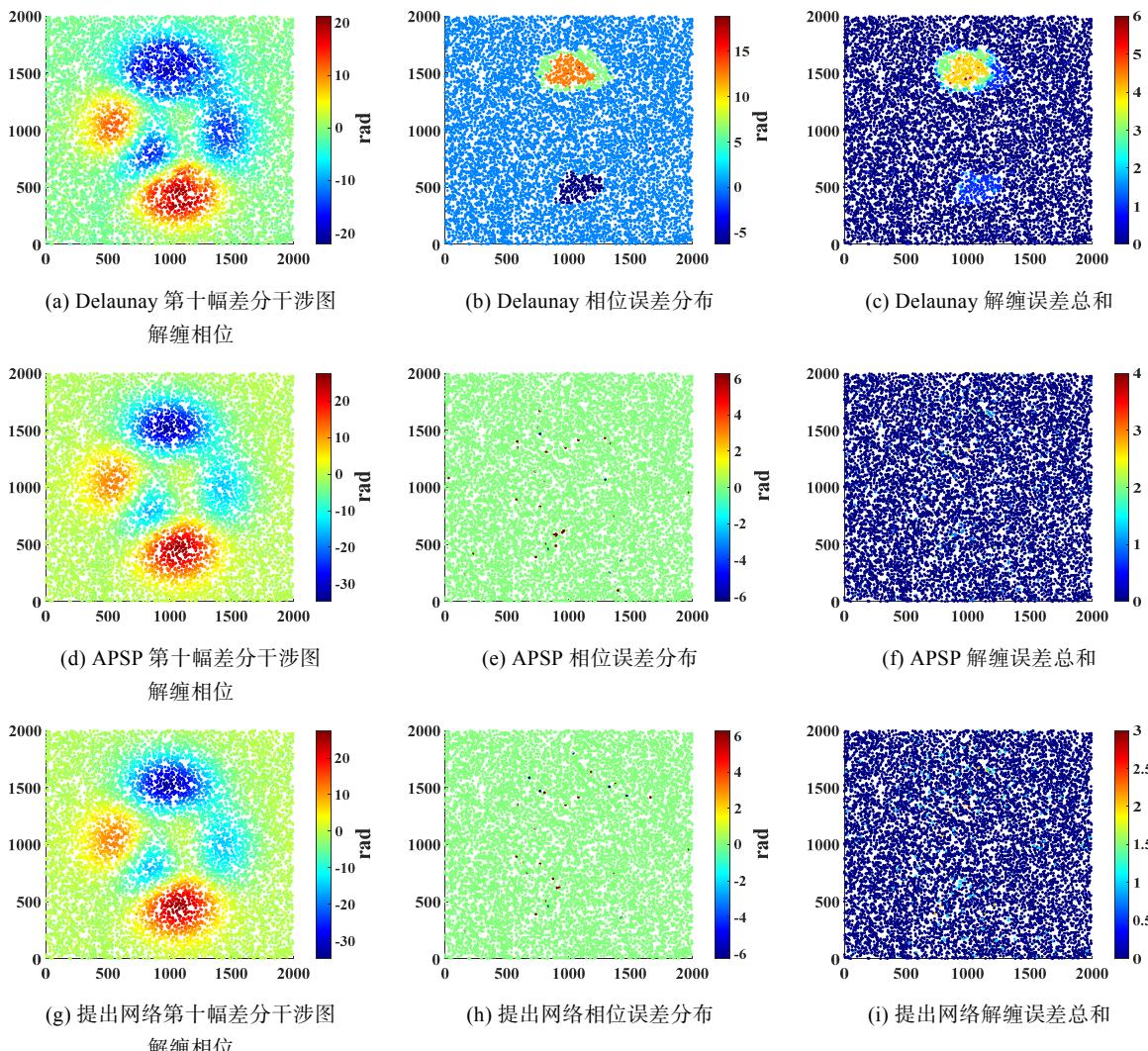


图 3 解缠结果对比

解缠边的可靠性评估中,较低的 MSE 和较高的时间相干性表明弧段具有较低的相位跃变可能性和较低的平均相位梯度,因此更为可靠。表 2 为不同网络连接的弧段统计结果,Delaunay 网络在拥有较高平均相位梯度及相位跃变可能性大的弧段数量较多,这影响其在处理大变形区域的相位解缠准确性。

表 2 不同网络弧段性能对比

	Delaunay 网络	APSP 网络	提出网络
MSE<0.9	6 719	7 911	9 073
0.9≤MSE<1.2	16 057	17 335	17 185
MSE≥1.2	7 190	5 050	4 030
$\gamma < 0.6$	11 486	8 379	6 046
$0.6 \leq \gamma < 0.8$	18 437	21 869	24 175
$\gamma \geq 0.8$	43	44	67
总边数	29 966	30 296	30 288

而采用时间相干性作为网络更新标准的 APSP 网络,在减少高平均相位梯度的弧段方面表现较为明显。相比之下,本文方法的优化网络在保持较低平均相位梯度和较小相位跃变可能性的弧段方面表现更为优越,表明其在满足相位连续性假设和提高网络整体可靠性方面的显著优势。

3 结论

本文提出了一种基于均方误差的网络优化时序 InSAR 相位解缠方法,通过采用最小二乘法计算各弧段的解缠相位,并利用均方误差对弧段进行评估和筛选,剔除误差较大的不稳定弧段。然后利用 Dijkstra 算法以时间相干性作为连接权重,重新构建剔除点与弧段顶点之间的连接,降低网络相位

(下转第 227 页)

基于 SCFDE-LDPC 无线模式的时间同步研究与实现

张 明^{1,2}, 段东辉^{1,2}, 芮缪苗^{1,2}

1. 北京无线电测量研究所, 北京 100854

2. 北京航天广通科技有限公司, 北京 100854

摘要: 时间同步是分布式系统的重要应用基础, 时间同步性能直接决定了分布式系统的功能和性能指标。介绍了一种基于单载波频域均衡 (SC-FDE) 调制解调方式和低密度奇偶校验码 (LDPC) 编译码的时间同步系统。系统采用无线的方式将本地及远端的设备进行时间同步, 达到远端设备基准与同步均与本地一致。仿真结果与实际使用效果表明方法有效稳定, 适合工程使用。

关键词: 单载波频域均衡; LDPC 编译码; 时间同步

文献标志码: A

中图分类号: TP23

Research and Implementation of Time Synchronization System Based on SCFDE-LDPC Communication Mode

ZHANG Ming^{1,2}, DUAN Donghui^{1,2}, RUI Miaomiao^{1,2}

1. Beijing Institute of Radio Measurement, Beijing 100854, China

2. Beijing Grantel Technology Co., Ltd., Beijing 100854, China

Abstract: Time synchronization is an important application foundation of distributed systems, and the performance of time synchronization directly determines the functionality and performance indicators of distributed systems. This paper introduces a type of time synchronization system based on single carrier frequency domain equalization (SC-FDE) modulation and demodulation method and low density parity check codes (LDPC). The system uses wireless means to synchronize time between local and remote devices, achieving consistency between remote device benchmarks and synchronization with local devices. The simulation results and actual application results indicate that the method is effective, stable, and suitable for engineering.

Key words: single carrier frequency domain equalization; low density parity check codes; time synchronization

1 引言

单载波频域均衡 (SC-FDE) 和正交频分复用 (OFDM) 技术都是为了克服符号间干扰提出的, 在实际传输的过程中, 由于多径的影响, 每条径的到达时间也不尽相同, 因此接收机无法确定符号的起始位置, 引起定时偏移; 定时误差使得接收机不能确定一帧的起始位置。此外, 在实际的应用中发射机和接收机的晶振频率不可能完全一致, 引起多普勒频移, 会造成接收端产生相位旋转, 甚至无法

正常解调。

SC-FDE 与 OFDM 的区别如图 1 所示^[1], SC-FDE 将原本在发送端的逆傅里叶变换 (IFFT) 移至接收端, 减轻了发送端的计算压力。低密度奇偶校验码 (LDPC) 编译码以其码率灵活、纠错性能优良、译码复杂度低、易于硬件实现等优点被广泛采用。

本文设计了基于无线通信的时统设备, 采用 SC-FDE 调制解调技术, LDPC 编译码技术, 保证

作者简介: 张明 (1990—), 男, 工学博士, 高级工程师, 研究方向为通信系统、数字信号处理。

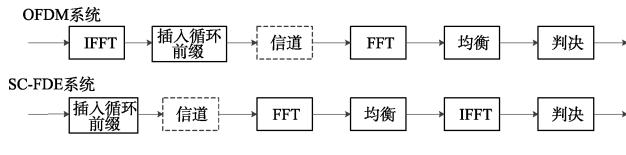


图 1 OFDM 与 SC-FDE 对比

设备的无线可靠传输与时统功能的可靠运行。将时统功能嵌入站间通信功能中，使设备具备传输业务数据的同时进行基准同步，并传输同步脉冲。

2 同步系统框图

同步系统的主要组成和功能框图如图 2 所示。同步系统包括本地端（基准端）与远端（恢复端），以本地端发射远端接收为例，介绍系统工作。本地端是时钟基准与同步基准的输入端，系统将这些模拟输入量通过数字量化转成数字数据，与通信业务数据复接，然后通过编码调制，从 DA 输出。再经过上变频等处理最终通过天线将信号发送出去。远端刚好相反，通过天线接收以及下变频等处理，将信号变为基带，再通过 AD 采集、解调解码、数据分接等处理，将同步信息提取，然后将同步基准信息恢复之后输出。

本地端采用系统提供的时钟基准（或内部产生的时钟基准）调制本地待发送信号，上变频等采用的基准也由此基准提供。此时调制的本地信号就带有了系统提供的时钟基准信息，远端在接收时将接收端的时钟信息与接收到的远端时钟信息做比对，远端就得到了本地端系统提供的时钟基准。通过解调解码得到的同步数据，计算收发时延，将收发时延补偿之后按照同步数据的控制字恢复本地端的同步基准，即为系统提供的同步基准。此时远端恢复的基准与系统提供的基准同时刻出现，实现远端的时间同步。

3 单载波频域均衡与 LDPC 编译码

3.1 单载波频域均衡技术

3.1.1 SC-FDE 原理

由于 OFDM 系统存在对定时误差、载频同步比较敏感和大的峰均比两点明显不足，如果采用单载波系统传输而保留 OFDM 系统对信号处理的方法，那么该系统在保持 OFDM 系统较好性能的同时避免了 OFDM 多载波系统的两大不足。为了克服多载波 OFDM 系统的缺点，Falconer 等提出了一种单载波频域均衡方案^[2]。在使用线性均衡的

SC-FDE 系统中，若傅里叶变换（FFT）运算的点数为 M ，则每个数据块所作乘法的次数为 $1b M$ ，与 OFDM 系统的复杂度相同。基本的单载波频域均衡的系统结构框图如图 3 所示。

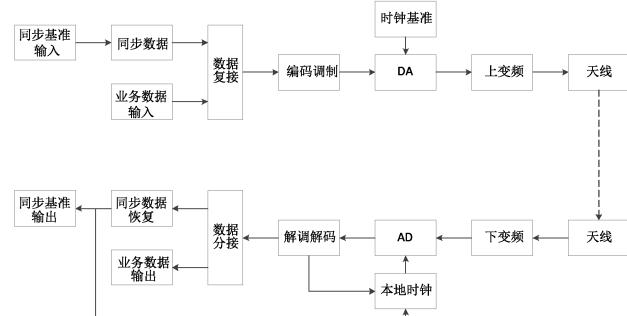


图 2 同步系统框图

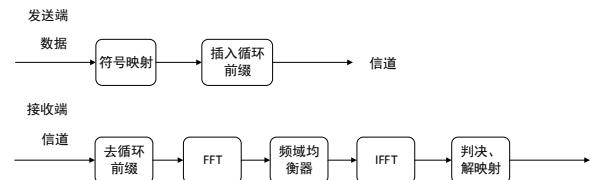


图 3 SC-FDE 系统结构框图

在发射端，数据符号经过映射后形成帧，并在每个数据帧之间插入保护间隔来最大限度地消除符号间干扰，保护间隔的长度一般要大于无线信道的最大时延扩展，这样一个符号的多径分量就不会对下一个数据帧造成干扰。

在接收端，对接收到的数据块进行删除循环前缀的操作，然后使用 FFT 将信号变换到频域，在频域经过均衡处理后，再通过 IFFT 操作变换回时域进行判决，得到重建的数据符号。

3.1.2 SC-FDE 数学表示

假设数据块 $\{a_k\}$ 的长度为 M ，传输速率为 $1/T$ ，每个数据块都带有循环前缀。接收信号为有 M 个取样的 $\{r_m\}$ ，滤波器的频域系数为 $\{W_l\}$ 。于是接收的复数信号可以表示为：

$$r_m = \sum_{k=0}^{M-1} h(mT - kT) a_k + n(mT), \quad (1)$$

$$m = 0, 1, \dots, M-1$$

其中， $h(\cdot)$ 是信道的冲激响应， $n(\cdot)$ 是加性噪声。由于存在循环前缀，可以假设 $\{a_k\}$ 具有周期性，对于任何整数 L ，有 $a_k = a_{k+LM}$ ，同时对于冲激响应也有 $h(mT) = h((m+LM)T)$ 。

在离散频域，式 (1) 可以表示为：

$$R_l = H_l A_l + V_l, l=0,1,\dots,M-1$$

式(2)中

$$H_l = \sum_{m=0}^{M-1} h(mT) \exp(-j2\pi) \frac{ml}{M} \quad (3)$$

$$A_l = \sum_{m=0}^{M-1} \exp(-j2\pi) \frac{ml}{M} a_m \quad (4)$$

$$V_l = \sum_{m=0}^{M-1} n(mT) \exp(-j2\pi) \frac{ml}{M} \quad (5)$$

经过频域均衡后，时间域的输出信号

$$Z_m = \frac{1}{M} \sum_{m=0}^{M-1} \exp(j\frac{2\pi}{M} ml) W_l R_l, \quad (6)$$

$$m = 0, 1, \dots, M-1$$

其中 $R_l = \frac{1}{M} \sum_{m=0}^{M-1} \exp(-j2\pi ml / M) r_m, l = 0, 1, \dots, M-1$ 即 R_l 是接收的 M 个信号 $\{r_m\}$ 的 FFT 变换。

采用最小均方误差(MMSE)准则，则滤波器的系数可以由下式得到：

$$W_l = \frac{H_l^*}{|H_l|^2 + \sigma^2} = \frac{H_l^*}{|H_l|^2 + 1/SNR}, \quad (7)$$

$$l = 0, 1, \dots, M-1$$

3.2 LDPC 编译码

3.2.1 LDPC 编码模块设计

Gallager 于 1962 年提出的 LDPC 码被称作低密度奇偶校验码，该名称来源于其校验矩阵是一种稀疏矩阵^[3]。LDPC 码是一种线性分组码，其译码性能可以逼近 Shannon 信道容量限^[4]，且编译码复杂度较低，可并行处理，因此被广泛应用于卫星通

信等工程中，尤其是高速率可靠传输中。

实际设计中，编码器实现结构示意图如图 4 所示。以 511 长度编码为例： A_i 为长度是 511 的循环移位寄存器， A_{i+j} 为 A_i 右移 j 位的结果， C 为长度是 511 位的寄存器。 $S_i \cdots S_{i+7}$ 为当前输入比特流。循环移位 A 右移 8 比特，向量 S 逐字节输入，进行上述同样的计算，直至完成全部元素的计算，得到校验信息。

3.2.2 LDPC 译码模块设计

LDPC 码的校验矩阵中的非零个数远远小于零元素的个数^[5]，是一类 (n, k) 线性分组码， n 为其码长， k 为其信息序列的长度，那么其校验矩阵的维数是 $(n-k) \times n$ 。

BP (置信传播) 译码算法^[6]的每一次迭代包括两步：校验节点处理和变量节点处理。下面将介绍该算法的主要步骤：

在 AWGN 信道下，采用 PSK 调制，将码字 c_i 映射为 $x_i = (-1)^{c_i}, i = 1, 2, \dots, n$ ， $L(P_i)$ 表示信道的初始消息， $L(P_i) = \log \left[\frac{\Pr\{x_i = 1 | y_i\}}{\Pr\{x_i = -1 | y_i\}} \right]$ ， $L(r_{ji})$ 表示校验传向变量的消息 $L(r_{ji}) = \log \left[\frac{r_{ji}(0)}{r_{ji}(1)} \right]$ ， $L(q_{ji})$ 表示变量节点传向校验节点的消息， $L(q_{ji}) = \log \left[\frac{q_{ji}(0)}{q_{ji}(1)} \right]$ ， $L(q_{ji})$ 表示变量节点收集到的所有消息。

信道初始化，概率 BP 算法的初始消息为：

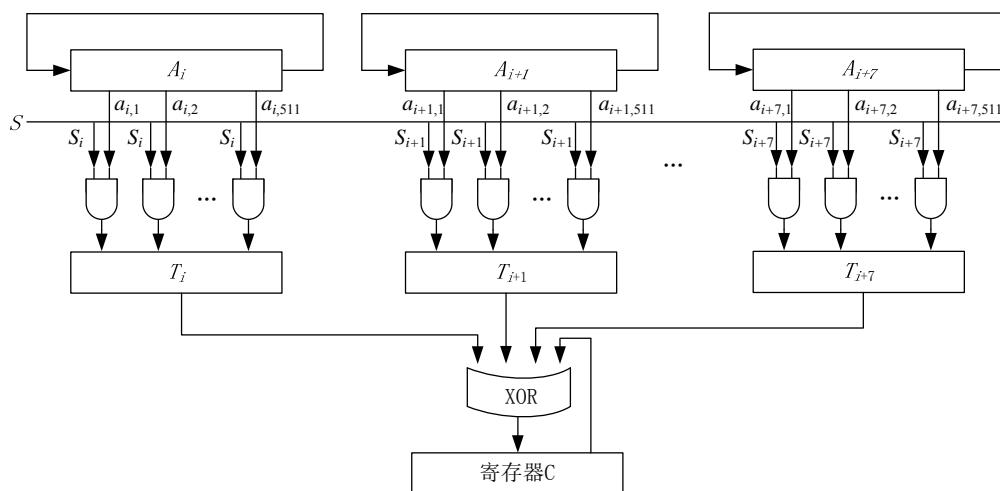


图 4 LDPC 编码示意图

$$\begin{aligned} q_{ij}^0(1) &= \Pr\{c_i = 1 | y_i\} = \Pr\{x_i = -1 | y_i\} = \\ &= 1 / (1 + e^{2y_i/\sigma^2}) \quad (8) \\ q_{ij}^0(0) &= \Pr\{c_i = 0 | y_i\} = \Pr\{x_i = 1 | y_i\} = \\ &= 1 / (1 + e^{-2y_i/\sigma^2}) \end{aligned}$$

得到对数似然比域 BP 算法的初始消息为：

$$L^0(q_{ji}) = \log \left(\frac{1 + e^{2y_i/\sigma^2}}{1 + e^{-2y_i/\sigma^2}} \right) = \log \left(e^{2y_i/\sigma^2} \right) = 2y_i / \sigma^2 \quad (9)$$

对数似然比域 BP 算法的译码步骤为：

初始化，对每一个变量节点 i 及与其相邻的校验节点进行初始化：

$$L^0(q_{ji}) = 2y_i / \sigma^2 \quad (10)$$

进行迭代处理，分为以下三个步骤：

步骤 1 对所有校验节点 j 数据更新：

$$L^{(l)}(r_{ji}) = 2 \tanh^{-1} \left(\prod_{i \in R_j} \tanh(L^{l-1}(r_{ji})) \right) \quad (11)$$

步骤 2 对所有变量节点 i 数据更新：

$$L^{(l)}(q_{ji}) = L(P_i) + \sum_{j \in C_i} L^{(l)}(r_{ji}) \quad (12)$$

步骤 3 进行译码硬判决：

$$L^{(l)}(q_i) = L(P_i) + \sum_{j \in C_i} L^{(l)}(r_{ji}) \quad (13)$$

如果 $L^{(l)}(q_i) > 0$ ，那么 $\hat{c}_j = 0$ ，否则 $\hat{c}_j = 1$ 。

判断迭代是否继续，如果将校验矩阵 H 与译码序列 \hat{c} 的转置相乘结果为零，即 $H\hat{c}^T = 0$ ，那么停止迭代。否则，如果迭代次数没有达到最大迭代次数，那么就又从步骤 1 继续迭代，反之停止迭代，输出译码判决结果。

4 远端恢复

收发使用的时钟频率相同相位不同，需要在发时钟中点时刻测量出收计时器的数值以及两个时钟的相位，如图 5 所示。

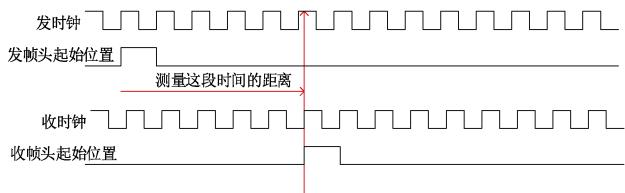


图 5 收发定时测量关系图

采用发时钟测量距离信息，首先用动态时钟生成一个周期为一帧的脉冲。使用该脉冲检测发帧头上升沿作为起始位置，开始滑动检测收帧头的上升沿作为终止位置。滑动的次数值就是发到收的距离。

以系统输入的同步脉冲为 $700 \mu s$ ，帧结构周期为 $200 \mu s$ 为例。本地端收到输入脉冲后，首先测量输入脉冲和本地端发计时器的位置。复现脉冲示意图如图 6 所示。

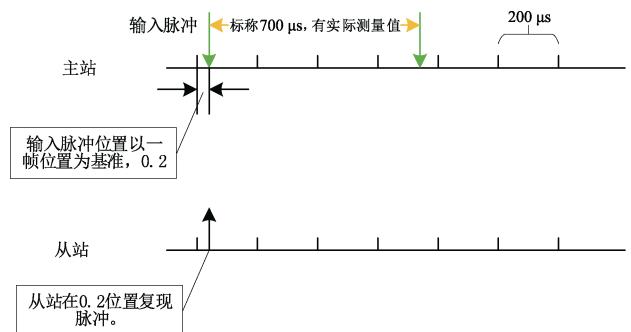


图 6 远端恢复输入脉冲示意图

脉冲到达时刻与所在帧结构位置决定了脉冲位置信息到达从站的延迟，这里将所有可能情况给出。假定在帧结构中点位置以前到达的脉冲信息可以在最近帧发送，在中点以后到达的脉冲信息在下一帧才能发送。

从图 7 上图可以看出，脉冲正常恢复，图 7 下图显示，当输入脉冲在中点位置以后出现时，传输该位置信息到从站后，会有一个输入脉冲到达，远端不能及时恢复脉冲。在这种情况下，需要远端具有计算出现脉冲位置并复现的功能。

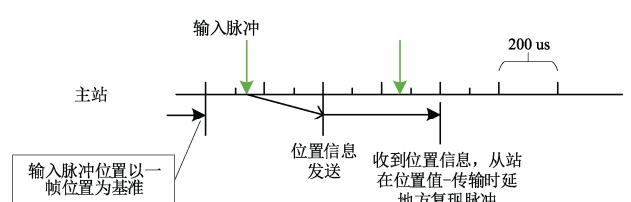
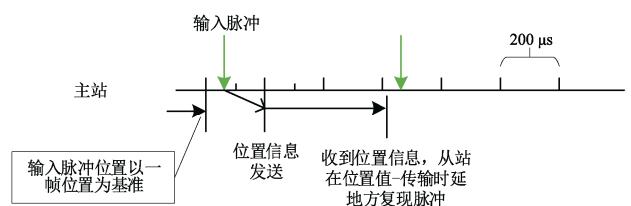


图 7 远端恢复同步脉冲示意图

因为 $200\ \mu s$ 的帧结构传输 $700\ \mu s$ 的脉冲会有相位模糊的情况，定义 $2\ ms$ 的超帧结构传输此 $700\ \mu s$ 脉冲，脉冲位置的计算和复现功能在此超帧上进行计算，超帧恢复 $700\ \mu s$ 同步脉冲如图 8 所示。远端收到同步脉冲控制字后在超帧之后开始恢复同步脉冲，黄色的为位置信息恢复的脉冲，红色的为计算守时恢复的脉冲，红黄相间的则可能是恢复的或者是守时得到的同步脉冲。



图 8 超帧与同步脉冲之间的关系

5 仿真与实验

5.1 脉冲传输恢复仿真

以 $700\ \mu s$ 脉冲传输为例，在本地模拟生成 $700\ \mu s$ 脉冲，然后输入到同步数据处理模块，经过调制解调以及编译码，按照时钟恢复原理，最终恢复出的 $700\ \mu s$ 如图 9 所示。



图 9 脉冲产生仿真图

从上到下依次为 $180\ MHz$ 时钟， $200\ \mu s$ 计数信号， $700\ \mu s$ 计数信号， $200\ \mu s$ 帧同步信号， $700\ \mu s$ 脉冲输出信号以及 $700\ \mu s$ 输出的计数器位置信号。

5.2 实际测试实验

采用模拟器接入 $700\ \mu s$ 脉冲信号，通过本地端以及远端的处理，在远端恢复 $700\ \mu s$ 脉冲信号。测试系统框图如图 10 所示，在实际测试时，模拟器具有固定的延迟，经过示波器校准，固定延时为 $321.5\ ns$ ，故实际输入至本地端的脉冲将比示波器显示的时刻滞后 $321.5\ ns$ 。本地端与远端建立时统之后，将远端恢复的脉冲接入示波器，与本地端输入的脉冲作比较，示波器结果如图 11 所示。

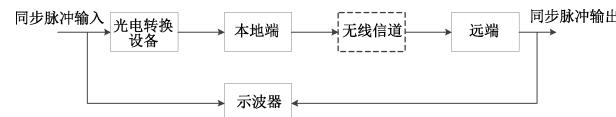


图 10 实际测试实验连接关系框图

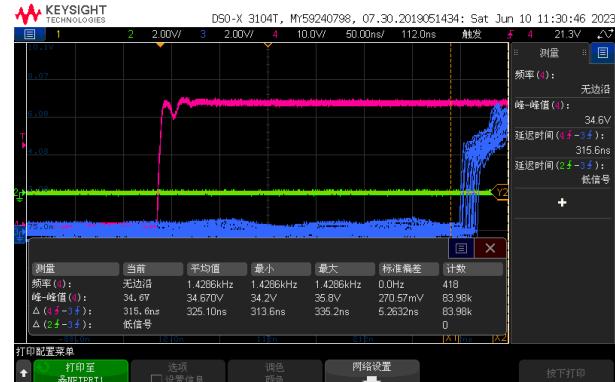


图 11 实际测试示波器截图

由示波器显示结果可得，脉冲在远端得以恢复，恢复抖动的标准差为 $5.2\ ns$ ，平均延时为 $325.1\ ns$ ，减去光电转换的固有延时误差，系统延时为 $3.6\ ns$ 。

6 结论

本文提出了一种基于 SCFDE-LDPC 通信方式的时频同步方法。SC-FDE 具有较低的峰均比，对射频通道的设计更加友好。LDPC 编译码属于高效的编译码方式，非常接近香农极限。采用此方式进行无线同步脉冲数据测量与恢复，拥有较强的鲁棒性。实验表明，脉冲恢复精度标准差为 $5.2\ ns$ ，具有良好的工程应用价值。

参考文献：

- [1] BAHAI A R S, Saltzberg B R, Ergen M. Multi-carrier digital communications: theory and applications of OFDM[M]. Springer, 2004.
- [2] FALCONER D, ARIYAVISITAKUL S L, BENYAMIN-SEYYAR A, et al. Frequency domain equalization for single carrier broadband wireless systems[J]. IEEE Communications Magazine 2002, 40(4): 58.
- [3] GALLAGER R. Low-density parity-check codes[J]. IRE Transactions on Information Theory, 1962, 8(1): 21-28.
- [4] CHEN J, DHOLAKIA A, ELEFTHERIOU E, et al. Reduced-complexity decoding of LDPC codes[J]. IEEE Trans. Comm, 2005, 53(8): 1288-1299.
- [5] 袁东风, 张海刚. LDPC 码理论和应用[M]. 北京: 人民邮电出版社, 2008.
- [6] GALLAGER R G. Low-density parity-check codes[J]. IRE Trans. Information Theory, 1962(1).

中轨搜救二代信号软件模拟器设计与实现

魏 嵘，张 波

北京航空航天大学 电子信息工程学院，北京 100191

摘要：中轨搜救系统是全球卫星搜救系统的重要组成部分，目前中轨搜救二代信标信号的接收机还处于从无到有的研发阶段，二代信号软件模拟器能够为接收机的测试提供大量信号数据，并且为系统问题的分析检测提供支持，因此二代信号软件模拟器的研发具有重要意义。基于二代信标信号的信号格式，设计了模拟器软件的整体架构以及工作流程，对软件各部分功能模块的算法进行了研究，具体研究包括：利用迭代法求解信号上下行链路传输时延；基于 BDS Klob 模型计算电离层时延，根据用户设置的位置信息匹配合适的对流层模型并计算对流层时延；计算接收时刻信号的码相位。最终实现了软件模拟器的基本功能，利用二代信标信号软件接收机对生成的数字中频信号进行了测试，检测分析了信号频谱以及信号的捕获、跟踪和解调结果。测试结果验证了生成信号的正确性，表明其可以用于接收机的研发测试。

关键词：中轨搜救；接收机测试；模拟器

文献标志码：A

中图分类号：TN911.72

Design and Implementation of Medium Earth Orbit Search and Rescue Second Generation Signal Software Simulator

WEI Wei, ZHANG Bo

School of Electronic and Information Engineering, Beihang University, Beijing 100191, China

Abstract: The medium earth orbit search and rescue system is an important component of the global satellite search and rescue system. Currently, the receiver of the second-generation beacon signal for medium earth orbit search and rescue is still in the development stage from scratch. The second-generation signal software simulator can provide a large amount of signal data for the testing of the receiver and support the analysis and detection of system problems. Therefore, the development of the second-generation signal software simulator is of great significance. This article is based on the signal format of second-generation beacon signals, and designs the overall architecture and workflow of the simulator software. The algorithms of each functional module of the software are studied, including: using iterative methods to solve the transmission delay of signals in the uplink and downlink; calculate ionospheric delay based on BDS Klob model, match appropriate tropospheric model according to user set location information, and calculate tropospheric delay; calculate the code phase of the received signal at the time of reception. The basic functions of the software simulator are ultimately implemented, and the generated digital intermediate frequency signal is tested using a second-generation beacon signal software receiver. The signal spectrum, as well as the results of signal capture, tracking, and demodulation, is detected and analyzed. The test results verify the correctness of the generated signal, indicating that it can be used for the development and testing of receivers.

Key words: medium earth orbit search and rescue; receiver test; simulator

基金项目：国家重点研发计划（2021YFB3901503）。

作者简介：魏嵒（2000—），男，硕士研究生，研究方向为卫星导航；张波（1972—），男，工学博士，副教授，研究方向为卫星导航、扩频通信。

全球卫星搜救系统(COSPAS-SARSAT)是由美国、前苏联、法国和加拿大四个国家联合开发的卫星搜救系统^[1],其任务在于提供准确可靠的求救信号和定位数据^[2]。截止到2011年,全球已有43个国家和组织成为C/S系统成员国^[3]。随着全球导航卫星系统(GNSS)的发展,其中轨(MEO)卫星的全球覆盖特性在搜救领域具有重要优势,全球卫星搜救系统开始逐步与GNSS开展合作,在其中轨卫星上搭载搜救载荷。该部分被称为中高度地球轨道搜索和救援系统(MEOSAR),它将补充现有的LEOSAR和GEOSAR系统^[4]。

目前MEOSAR系统已被确定为国际搜救服务开展的主要研究方向,将其实现过程分为多个阶段^[5]。MEOSAR系统将进一步增强全球卫星搜救系统的功能,成为其中的重要组成部分。北斗导航系统是第三个已经投入实际运行的卫星导航系统^[6],它的应用产生了显著的经济效益和社会效益^[7]。北斗系统的加入也促进其在国际民用航空、航运等多个领域的发展。北斗系统将凭借其优越的性能为全球遇险信标的进行精确定位,为国际搜救行动提供服务,这也将有力地推动北斗全球化发展,使得我国在搜救卫星领域的影响力越来越大^[8]。

随着导航模拟器技术不断发展,目前市场上已有GSS、NCS等多种系列模拟器^[9-10],而二代搜救信号模拟器软件的研发还处于起步阶段。其主要目的在于为地面站接收机研制和系统集成提供测试验证手段。二代中轨搜救信号接收机是中轨搜救系统地面站建设的重要部分,目前仍处于建设和发展阶段,接收机的研发过程中需要大量的仿真数据用于测试与验证。二代搜救信号模拟器软件能够根据北斗二代搜救信号的特性以及用户的需求模拟信号的传输过程,最终生成所需的仿真信号,为软件接收机的测试提供可靠的数据支持。此外,北斗中轨搜救系统运行过程中产生问题时,利用二代搜救信号模拟器软件可以准确地还原信标信号通过上行链路传输至搜救载荷,再由载荷转发至地面站的全部过程,对系统问题分析和检测具有重要的意义。

1 中轨搜救二代信标信号结构

1.1 二代信标信号调制方式

中轨搜救系统第二代遇险信标信号采用的调

制方式是直接序列扩频-偏移四相相移键控(DSSS-Offset-QPSK, DSSS-OQPSK)。OQPSK是以四相移相键控(QPSK)作为基础,通过改进得到。DSSS-OQPSK调制信号可由式(1)表示:

$$S(t) = AI(t)[\sum a_n g_T(t - nT_s)] \cos \omega_c t + \\ AQ(t)[\sum b_n g_T(t - nT_s - T_c / 2)] \sin \omega_c t \quad (1)$$

DSSS-OQPSK调制的二代信标信号同相支路和正交支路采用不同的截短的m序列扩频码,码长为38 400。扩频伪码通过线性反馈移位寄存器(LFSR)生成,其中正交支路(Q)和同相支路(I)的伪码生成器使用相同的生成多项式 $G(x) = X^{22} + X^{18} + 1$ 以及不同的LFSR初相。

DSSS-OQPSK信号调制器结构如图1所示,首先将300 bit电文数据进行串并转换为奇序列和偶序列,奇序列通过同相支路(I)传输,偶序列通过正交支路(Q)传输,其中每一个信息比特位对应256个扩频码片;将I、Q两路150 bit的基带数据流分别与相应支路的38 400 bit扩频伪码序列进行模二加,得到扩频调制后的数据,将扩频后的Q支路数据序列延迟半个伪码的码片周期;将I、Q两路扩频信号经过低通滤波器进行成型滤波,之后将两路信号分别与频率为406.05 MHz的正弦和余弦载波进行调制;将两路信号合并,经过低噪放和滤波器,最后通过天线发送。

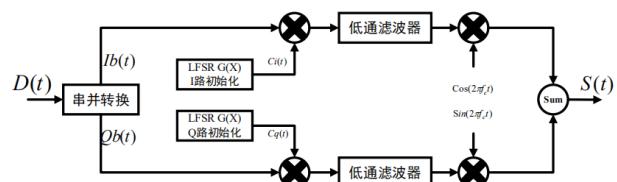


图1 DSSS-OQPSK信号调制器

1.2 二代信标信号电文格式

1.2.1 帧结构

二代搜救信号为突发信号,信号数据长度为300 bit,持续时间为1 s,由前导码、有用信息段、纠错码三部分组成。其中,前导码长度为50 bit,不包含有用信息,数据序列全为“0”;信息段长度为202 bit,主要提供了遇险信号的位置等信息;纠错段长度为48 bit,包含用于接收端纠错的比特,采用BCH码。二代搜救信号的突发帧结构如图2所示。

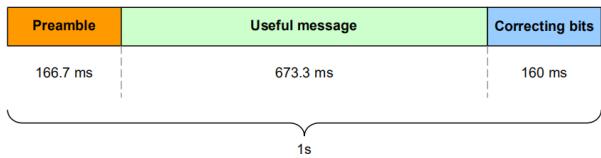


图 2 二代搜救信号的突发帧结构

1.2.2 数据校验位

二代信标信号中使用的纠错代码是 BCH(255,207)代码。该代码由 207 位数据和 48 位 6 重纠错码组成。该代码用于检测和纠正 250 位数据比特(二代搜救信标信号的有用信息段和纠错码部分)中的 6 个错误。

为了正确生成 BCH(255,207)代码, 需要对数据部分进行填充。因此, 在 202 位数据之前添加 5 个“0”数据比特, 形成代码的 207 位模式。对于 BCH(255,207)码, 定义一个生成多项式如式:

$$g(X) = \text{LCM}(m_1(X), m_3(X), m_5(X), m_7(X), m_9(X), m_{11}(X)) \quad (2)$$

其中, LCM 为最小公倍数, 式中 $m(x)$ 值分别为:

$$\begin{aligned} m_1(X) &= X^8 + X^4 + X^3 + X^2 + 1 \\ m_3(X) &= X^8 + X^6 + X^5 + X^4 + X^2 + 1 \\ m_5(X) &= X^8 + X^7 + X^6 + X^5 + X^4 + X + 1 \\ m_7(X) &= X^8 + X^6 + X^5 + X^3 + 1 \\ m_9(X) &= X^8 + X^7 + X^5 + X^4 + X^3 + X^2 + 1 \\ m_{11}(X) &= X^8 + X^7 + X^6 + X^5 + X^2 + X + 1 \end{aligned} \quad (3)$$

202 位数据的多项式如下:

$$m(X) = b_1 X^{201} + b_2 X^{200} + \dots + b_{201} X + b_{202} \quad (4)$$

最后将代码扩展至 207 位, 用 5 个“0”数据比特填充前 5 位。将扩展后的 207 位二进制字符串除以 $g(X)$, 计算得到的余数即所需的纠错码, 以上就是二代信标信号的 BCH(255,207)编码过程。

2 二代信号模拟器方案设计

2.1 模拟器总体设计

模拟器主要实现了北斗搜救信号测试仿真系统中载荷仿真模块的功能, 模拟器软件系统能够根据设置的场景参数, 模拟信标信号经过搜救载荷接收并转发至地面站的信号传输过程, 计算出所选取的卫星状态参数, 生成四颗卫星对应数字中频信号。模拟器软件系统的整体总体设计如图 3 所示。

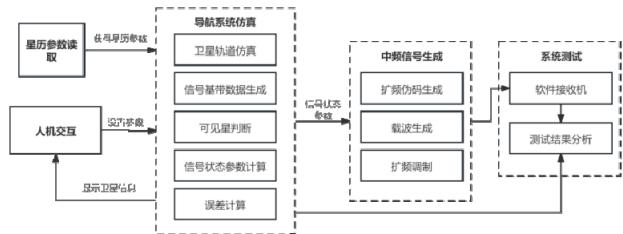


图 3 模拟器软件系统的总体设计

2.2 模拟器功能模块

北斗中轨二代搜救信号模拟器软件采用模块化的设计方式。该设计方式使软件的整体逻辑结构更加清晰, 同时有利于对软件各部分进行测试和升级, 而不需要对程序进行大幅度的改动^[11]。

将软件按照大致功能划分为四个模块, 分别为: 人机交互模块、功能模块、参数计算模块和中频信号生成模块。各模块再向下细分为多个子模块。各模块之间的关系如图 4 所示。

人机交互模块主要实现了可视化界面, 用户能够通过界面设置参数(信标位置、接收机位置、仿真时间、搜救载荷转发时延等), 同时将仿真结果显示在界面上; 功能模块主要包括伪码生成模块、信号电文生成模块和星历文件读取模块; 参数计算模块主要包括时间坐标转换模块、卫星位置计算模块、可见星判断模块、信号状态参数计算模块、误差计算模块等功能模块, 主要作用为根据用户设置的参数计算得到各个可见星的信号状态参数(多普勒频率、码相位、信号传输时延等); 中频信号生成模块主要包括载波和伪码生成模块和扩频调制模块等, 主要功能为接收系统仿真部分计算得到的参数, 根据二代搜救信号的体制以及调制方式对基带数据进行扩频调制, 生成数字中频信号并以文件形式保存。为保证软件正常运行, 需要确定各模块的执行顺序^[12]。

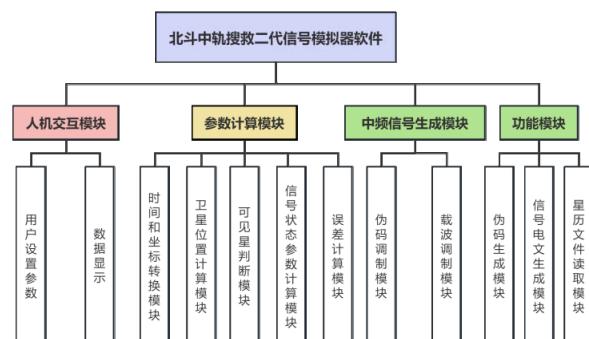


图 4 模拟器软件各模块

2.3 软件工作流程

模拟器软件的整体工作流程图如图 5 所示。首先获取用户设置的仿真参数，读取 RINEX 格式的星历文件获取星历参数。经过参数计算模块得到信号发送时刻，计算卫星位置，判断卫星的可见性。计算信号传输过程中的主要误差，计算上下行信号的多普勒频率、伪码相位和载波相位等参数。最后利用生成的扩频伪码、载波和电文数据进行扩频调制，量化生成数字中频信号，并以文件形式保存。

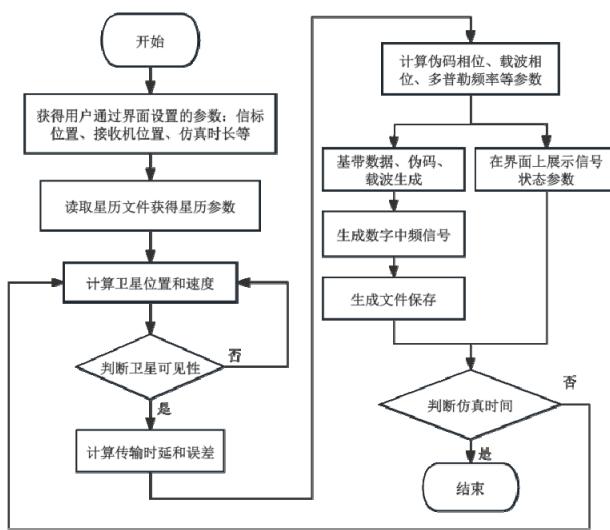


图 5 模拟器软件工作流程

3 主要功能模块设计

3.1 信号传输时延计算模块

从信号播发到接收机全过程的时延包括上行链路传输时延、下行链路传输时延和载荷转发时延。利用迭代法计算信号传输时间 τ 的原理如下，以下行链路传输过程为例：

已知卫星接收信号时刻 t_{r1} ，设卫星转发信号的时刻为 t_{l1} ，首先确定传输时延的初始值 $\Delta t_R = 0.075\text{ s}$ ；

计算卫星转发信标信号的时刻 t_{l1} ，接收时刻已知，转发时刻 $t_{l1} = t_{r1} - \Delta t_R$ ；

计算转发时刻 t_{l1} 卫星的坐标 (X_s, Y_s, Z_s) ；

计算转发时刻卫星和当前接收机位置的几何距离 R ，接收机位置已知，几何距离 $R = \sqrt{(X_s - X_r)^2 + (Y_s - Y_r)^2 + (Z_s - Z_r)^2}$ ；

计算几何传输时延 $\Delta t = R / C$ ，其中 C 为光速；

计算 Δt 和 Δt_R 之间的误差 δ ， $\delta = |\Delta t - \Delta t_R|$ ；

将 δ 与门限值 ϵ 进行比较，其中门限值 ϵ 设置为 10^{-12} ，此时的伪距误差约为 0.3 厘米，如果 $\epsilon > \delta$ ，则此时的几何传播时延 Δt 为生成信号时使用的数值，如果 $\epsilon < \delta$ ，用计算得到的几何传播时延 Δt 替换几何传播时延的初值，即 $\Delta t_R = \Delta t$ ，之后重复上述过程。上行链路传输时延的求解过程同理。信号传输时延计算的流程如图 6 所示。

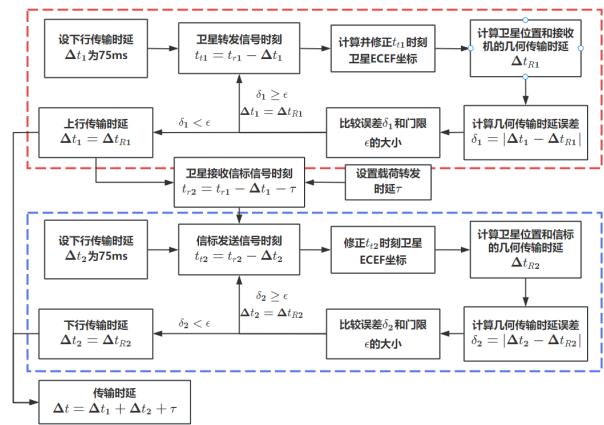


图 6 信号传输时延计算流程图

3.2 误差计算模块

3.2.1 电离层误差

电离层模型采用 BDS Klob 模型，该模型在 Klobuchar 模型的基础上进行了优化和改进，在中国区域相比 GPS 的 Klobuchar 模型，定位精度提高了约 10%^[13]。它引入了一个电离层时延修正精度 redial，根据不同的适用条件来调整修正精度参数可以使模型实现更精确的误差修正，从而提高模型的适用性。

在高纬度地区，信号在电离层中传播的路径相对较长，电离层对传输时延造成的影响更加明显，时延修正精度参数较低。而在低纬度地区，信号在电离层中传播的路径相对较短，时延修正精度参数较高。此外，昼夜和月份也会对时延修正精度参数造成影响，所以，在计算电离层误差时需要根据纬度、昼夜和月份的不同调整时延修正精度参数，从而得到更准确的计算结果。

3.2.2 对流层误差

常用的对流层模型包括 Black 模型、Saastamoinen 模型和 Hopfield 模型等。各模型适用的湿度条件不同，Black 模型适用于高湿度环境，Saastamoinen

模型适用于中湿度环境, Hopfield 模型适用于低湿度环境^[14]。

对流层延迟的计算流程如图 7 所示, 根据用户输入的高度数据, 利用标准大气模型计算得到该位置的温度和气压; 根据用户输入的经纬度数据, 匹配得到该地区的降水量和湿度, 从而选择适用的对流层模型; 最终根据所选模型以及参数计算得到对流层延迟。通过以上流程有效提高了模型的实用性和可靠性^[15]。

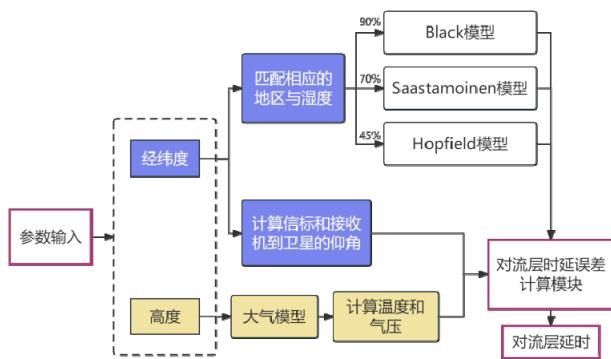


图 7 对流层延迟计算流程图

3.3 码相位计算模块

设置信标开始发送信号的时刻为 t_0 , 则在 t_r 时刻接收机接收到的码相位和信号比特位置计算过程如下:

(1) 判断接收机在 t_r 时刻是否接收到信号。假设在 t_r 时刻有信号到达接收机, 则信号的发射时刻为:

$$t_t = t_r - \Delta t - \Delta t_{\text{trop}} - \Delta t_{\text{ion}} + \Delta t_{\text{sc}} \quad (5)$$

其中, t_r 表示接收机接收到信号的时刻, Δt 为信号从信标到接收机的传输时延, Δt_{trop} 表示对流层

延迟, Δt_{ion} 表示电离层延迟, Δt_{sc} 为星钟误差。

(2) 计算发射时刻 t_t 与信标开始发送信号的时刻 t_0 的时间差:

$$\Delta \tau = t_t - t_0 \quad (6)$$

(3) 若 $\Delta \tau < 0$, 说明此时信号还未发射, t_r 时刻接收机没有信号到达。

(4) 若 $\Delta \tau \geq 0$, 则根据 $\Delta \tau$ 的长度进行如下处理, 其中 $\text{fmod}()$ 表示取余运算。

$$\tau = \begin{cases} \text{fmod}(\Delta \tau, 5), 0 \leq \Delta \tau < 30 \\ \text{fmod}(\Delta \tau, 30), 0 \leq \Delta \tau < 30 \\ \text{fmod}(\Delta \tau, 120), 0 \leq \Delta \tau < 30 \end{cases} \quad (7)$$

若 $0 \leq \Delta \tau \leq 1$, 则 t_r 时刻接收机有接收到信号, 否则接收机没有接收到信号。

(5) 计算伪码相位 T_{wm} 和电文数据位 T_d , 其中 $\text{INT}[]$ 为向下取整函数。

$$T_{\text{wm}} = \text{INT}[\tau \times 38400] \quad (8)$$

$$T_d = \text{INT}[\tau \times 150] \quad (9)$$

4 测试与验证

4.1 生成信号测试

首先对生成信号进行测试。设置仿真时刻为 2024 年 6 月 28 日 11:40:00, 信标位置为东经 123.6°, 北纬 27.1°, 高度 0 m, 接收机位置为东经 116.3°, 北纬 39.9°, 高度 0 m, 卫星屏蔽角为 5°。模拟器软件的人机交互界面的参数设置与仿真得到的可见星的信号状态参数结果如图 8 所示。结果显示可见星数目为 4, 分别为 12 号、34 号、35 号和 44 号星, 同时界面展示了卫星位置、多普勒频率和传输时延等信息。

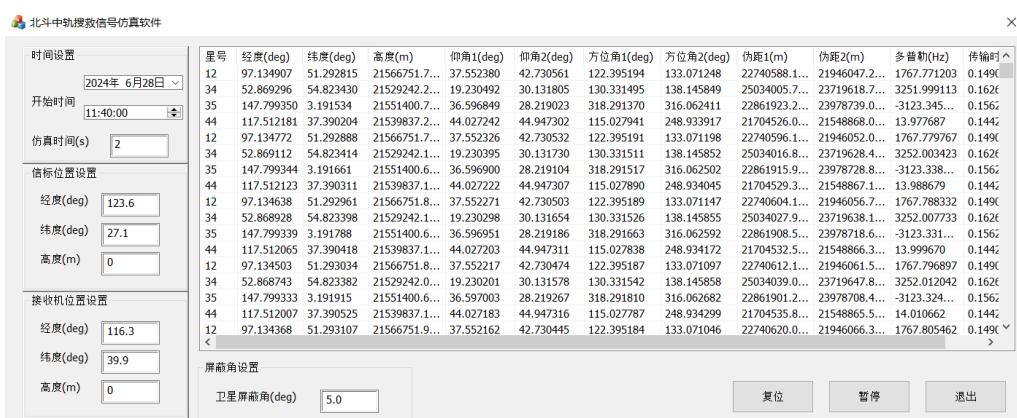


图 8 软件仿真结果

取 12 号星生成数字中频信号。设置信号载波频率为 62 270 Hz, 采样率为 384 000 Hz, 载荷转发时延为 0.1 s, 信标发送延迟为 0。设置信号电文前 50 bit 为 0, 之后 202 bit 开始取 50 bit 为 1、50 bit 为 0 依次交替, 最后 48 bit 为纠错位。生成数字中频信号如图 9 所示。

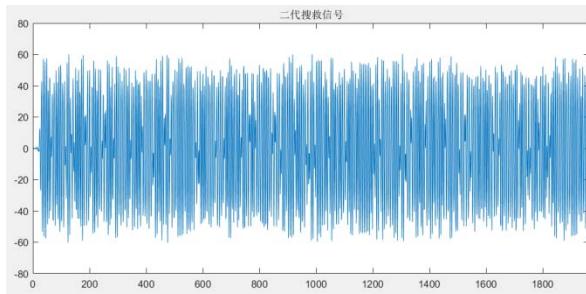


图 9 二代信标信号

信号的频谱如图 10 所示。由频谱图像可知, 信号频率分布主要集中于一定范围, 符合 DSSS-QQPSK 调制方式的频谱结果特点, 证明了生成信号满足扩频调制的频谱特性。

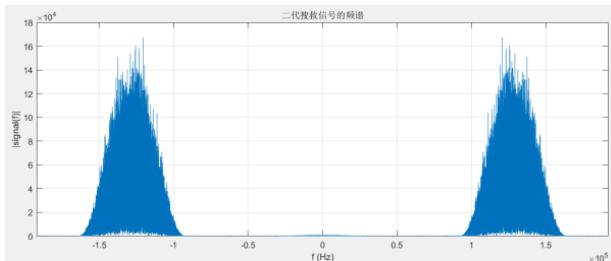


图 10 二代信标信号频谱

4.2 软件接收机测试

利用二代搜救信号软件接收机对生成的数字中频信号进行测试。I 路伪码捕获结果如图 11 所示, 从捕获结果可以看出自相关峰大于设置的捕获门限 0.8, 证明成功捕获, 下一步将进行载波和伪

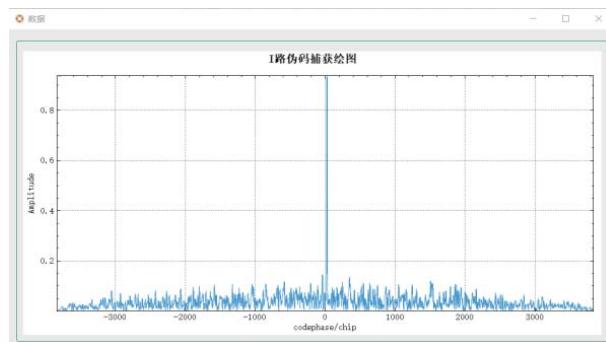


图 11 I 路伪码捕获结果

码跟踪处理。

锁相环频率误差跟踪结果如图 12 所示, 由于相较于载波, 信号的多普勒频率变化较小, 载波频差基本保持不变, 所以载波频率跟踪结果显示信号频差的变化幅度较小, 基本稳定于一个定值。

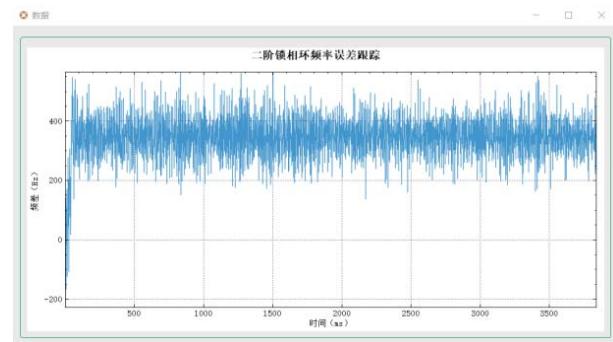


图 12 锁相环频率误差跟踪结果

锁相环相位误差跟踪结果如图 13 所示, 相位误差跟踪结果在初始时刻进行了快速振荡, 之后逐渐趋于稳定状态, 最终在 0 的附近小幅度振荡。以上跟踪结果表明了信号跟踪成功, 实现了频率和码相位的同步。

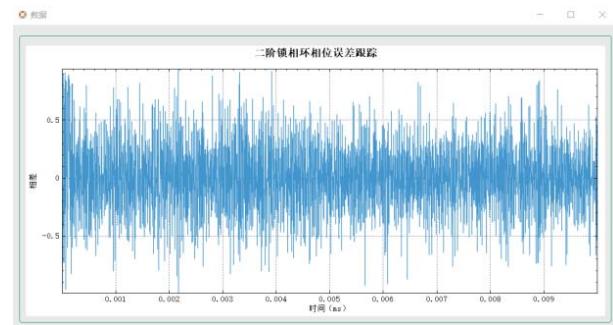


图 13 锁相环相位误差跟踪结果

I 路和 Q 路解调得到的数据结果如图 14、15 所示, I、Q 两路解调结果均为 0、1 交替的数据, 将两路数据合路后与设置的基带数据进行分析, 经过

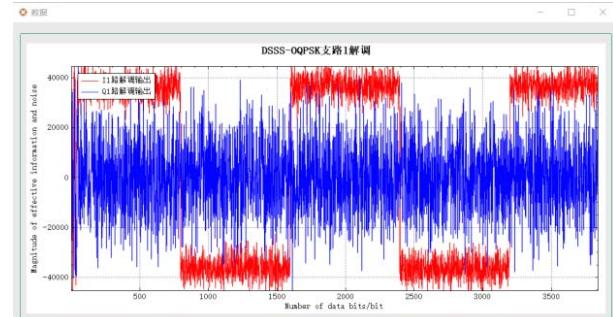


图 14 支路 I 解调结果

(下转第 254 页)

有限状态机在网络设备信息采集技术中的应用与研究

赵 鹏，熊宗炬，张万彬，张小锋，吴树盛

中国西昌卫星发射中心，四川 西昌 615000

摘要：有限状态机通过一组离散的状态集合和在这些状态之间的转移规则来描述系统的行为。每个状态代表系统在特定时间点的行为或条件，而状态之间的转移由输入事件或条件触发。现有网管系统存在设备异构性、配置管理不一致性及设备批处理方面存在局限。提出一种基于有限状态机的解决方案，实现异构网络设备的统一管理与配置，并提供了核心框架代码，展示了其在网络设备登录和配置管理过程中的应用，能够适应性地处理网络设备的异构性，并为网络设备的统一监控和维护提供支持。

关键词：有限状态机；网络设备；状态迁移；集中配置

文献标志码：A

中图分类号：TP393

Network Device Information Collection Technology Based on Finite State Machine

ZHAO Peng, XIONG Zongju, ZHANG Wanbin, ZHANG Xiaofeng, WU Shusheng

Xichang Satellite Launch Center, Xichang, Sichuan 615000, China

Abstract: Finite state machine describes the behavior of a system through a set of discrete states and transition rules between these states. Each state represents the behavior or condition of the system at a specific point in time, and the transitions between states are triggered by input events or conditions. Existing network management systems have limitations in terms of device heterogeneity, configuration management inconsistency, and batch processing of devices. This paper proposes a solution based on the finite state machine, which achieves unified management and configuration of heterogeneous network devices, and provides the core framework code. It demonstrates its application in the process of network device login and configuration management, capable of adaptively handling the heterogeneity of network devices, and providing support for the unified monitoring and maintenance of network devices.

Key words: finite state machine; network devices; state transition; centralized configuration

随着计算机网络技术的迅猛发展，网络架构的规模不断扩大，网络中的设备数量显著增加，一些单位的网络设备已经达到了数千台的量级。此外，交换机作为网络中的核心组件，其型号和规格也呈现出多样化的趋势，为网络的管理和维护带来了新的挑战。这些交换机型号不同、管理方式不同、认证方式不同，甚至配置命令也大同小异^[1]。如何能够有效对全网交换机进行管理，始终是一个难题。

现有的网管系统，对设备型号一致、登录方式一致的网络设备可以实现很好的管理。然而。然而，设备类型的多样化，网络管理系统在实现有效监控和管理方面遇了一系列挑战。首先，网络设备

的异构性导致了管理协议和策略的不一致性，增加了配置和兼容性的复杂性。其次，随着设备数量的增加，网络流量的监控和分析变得更加困难，需要更高效的数据处理和分析算法来应对大数据量。本文基于有限状态机，结合 telnet、ssh、snmp、http 等技术，实现了不同类型、不同型号网络设备的信息采集和集中配置管理^[2]。

1 有限状态机

有限状态机（finite state machine, FSM）是一种抽象的计算模型，它通过定义一组离散的状态集合以及在这些状态之间的转移来模拟系统的行为^[3]。在有限状态机中，每个状态都代表了系统

在某一特定时间点的特定行为或条件。状态之间的转移是由输入事件或条件触发的，这些事件或条件被称为触发器。

一个有限状态机 M 可以表示为五元组 $(Q, \Sigma, \delta, q_0, F)$ ：

状态集合 Q : 一组预定义的状态，每个状态都具有唯一的标识符。每个状态 q 属于 Q 。

输入集合 Σ : 所有可能输入的集合，一组可能影响状态转移的输入事件。每个输入 σ 属于 Σ 。

转移函数 $\delta: Q \times \Sigma \rightarrow Q$: 一个函数，根据当前状态和输入事件来确定下一个状态。

初始状态 q_0 : 有限状态机开始时的状态， q_0 属于 Q 。

终止状态集合 F : 一组特定的状态，当有限状态机达到这些状态时，它被认为是已经完成了它的任务。每个终止状态 f 属于 F ，且 $F \subseteq Q^{[4]}$ 。

有限状态机的操作可以描述为：在给定的初始状态下，系统接收输入事件，并根据转移函数转移到新的状态^[5]。这个过程会持续进行，直到系统达到一个终止状态，此时有限状态机完成其任务。

有限状态机广泛应用于计算机科学和工程领域，包括但不限于编程语言的词法分析、网络协议的实现、数字电路设计以及各种自动化控制系统。它们提供了一种简洁且系统化的方法来设计和理解复杂系统的行为。

2 核心框架代码

在实际编程实现状态机时，通常涉及使用条件语句（比如 if-else）或者查表来管理状态转移逻辑。有限状态机提供了一种清晰和结构化的方法处理复杂的条件和行为，使得系统的设计和调试更加容易^[6]。此外，状态机模型也有助于确保系统在任何时候都处于已知的状态之一，从而提高系统的可靠性和可预测性。

本文用 python 实现了一个简单的有限状态机，定义了一个名为 StateMachine 的类，该类提供了创建有限状态机、添加状态、设置初始状态、运行状态机以及处理状态转移的基本功能，流程图如图 1 所示。

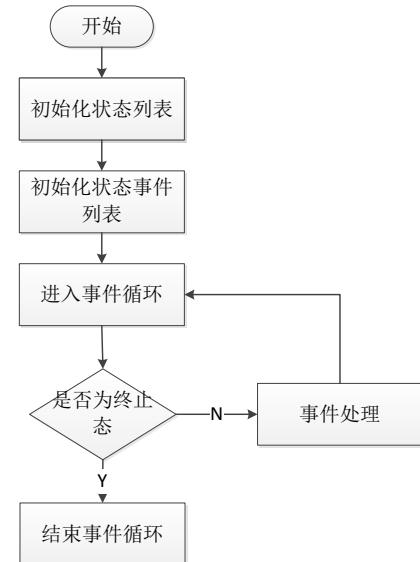


图 1 有限状态机实现流程图

核心代码如下所示：

```

class StateMachine:
    def __init__(self):
        self.handlers: Dict[str, Handler] = {}
        self.start_state: str = None
        self.end_states: List[str] = []
    def add_state(self, state: str, handler: Handler):
        state = state.upper()
        self.handlers[state] = handler
        self.end_states.append(state)
    def set_start(self, state: str):
        self.start_state = state.upper()
    def run(self, cargo: Dict[str, Any]):
        try:
            handler = self.handlers[self.start_state]
        except:
            raise Exception('call .set_start() before .run()')
        if not self.end_states:
            raise Exception('one state must be an end_state')
        while True:
            (new_state, cargo) = handler(cargo)
            if new_state.upper() in self.end_states:
                handler = self.handlers[new_state.upper()]
                if handler:
                    handler(cargo)
                    break
                else:
                    handler = self.handlers[new_state.upper()]
            return cargo
  
```

其中：

`handlers`: 一个字典，用于存储状态转移函数，键是状态名，值是对应的处理函数。

`start_state`: 初始状态，开始运行状态机时所在的状态。

`end_states`: 一个列表，包含所有终止状态的名称。

`cargo` 参数: 状态机运行时携带的数据，可以

是任何类型，取决于状态转移函数的处理需求，本文使用了字典数据类型携带数据。当达到终止状态后，调用终止状态的处理函数，然后退出循环，返回 cargo 数据^[7]。

3 全网网络设备遍历

3.1 有限状态机设计

考虑到网络设备在登录协议、认证机制以及凭证（用户名和密码）方面存在显著的异质性，该机制能够适应性地处理各种差异性，确保对每个设备的访问和数据收集过程的一致性和有效性^[8]。

具体而言，有限状态机通过预定义的状态集合和转移函数，对网络设备的登录过程进行建模。状态集合涵盖了从初始化、身份验证到登录成功或失败的各种可能状态。转移函数则根据设备的特定登录要求和认证结果，动态调整状态转移的路径。此外，有限状态机还包括了处理异常状态和重试机制的策略，以应对登录过程中可能出现的认证失败或网络错误等问题。

根据网络设备的不同登录方式、用户名、密码组合方式不同、配置信息不同等，超过了 100 多种可能组合，需要根据网络设备的返回的信息，确定设备所处的登录状态，根据事先定义好的事件进行处理。根据实验环境，定义了如图 2 的网络设备遍历登录有限状态机。

3.2 信息获取流程

考虑到不同网络设备可能采用多种远程登录协议，本机制首先尝试利用 Telnet 协议对设备进行访问。在此基础上，根据交换机的响应反馈，系统动态判断并选择适当的登录协议，即 Telnet 或

SSH，以适应交换机的特定需求。

具体而言，系统初始化时，通过 ssh 协议向目标设备发送登录请求。随后，系统捕获并解析交换机的响应信息，以确定适用的登录协议。若交换机返回特定的系统标识符“<sysname>”，表明设备已准备就绪并可进入配置模式 “[sysname]”。此时，系统认为该设备已成功登录，并且满足远程管理的要求。

成功登录后，系统将相关的登录凭证、协议类型及设备响应信息记录至本地 SQLite 数据库中，以实现对网络设备登录信息的持久化存储和管理^[9]。具体流程如图 3 所示。

4 全网网络设备信息获取和配置

4.1 有限状态机设计

网络设备信息获取，根据上文，已经获取了网络设备的登录方式。下一步可以登录到网络设备获取配置信息、运行信息，以及对网络设备进行配置。获取网络设备配置信息的有限状态机如图 4 所示。

为了获取网络设备的登录信息后，需要进一步采集设备的配置信息以及集中批量对网络设备进行管理。首先，通过已建立的认证机制实现对交换机的登录，随后输入特定的命令集以收集所需数据。

4.2 配置流程

在设备配置方面，实现了两种策略：单设备配置和全网设备的批量配置。在单设备配置过程中，系统首先验证设备的可达性，然后通过远程登录对目标设备应用配置命令。而在全网设备的批量配置策略中，系统将首先测试网络中所有设备的连通性，确保每个设备均处于可访问状态。继而，系统

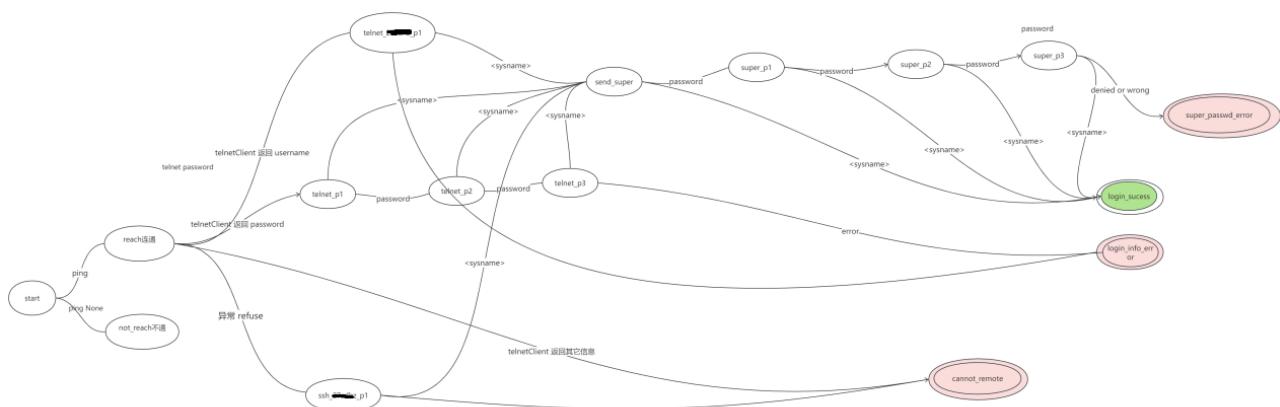


图 2 网络设备列表遍历状态机

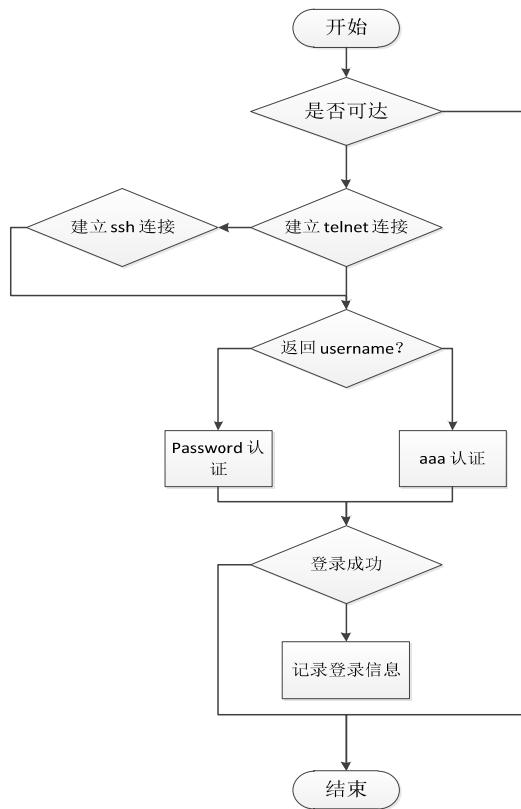


图3 可登录遍历流程图

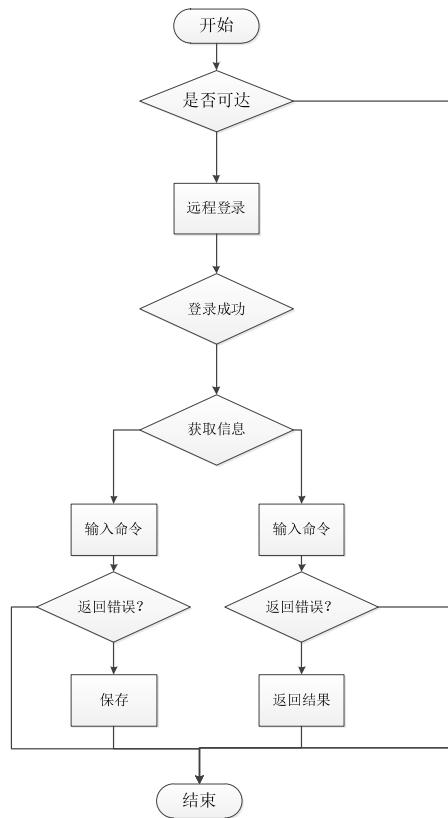


图5 网络设备信息获取和配置流程图

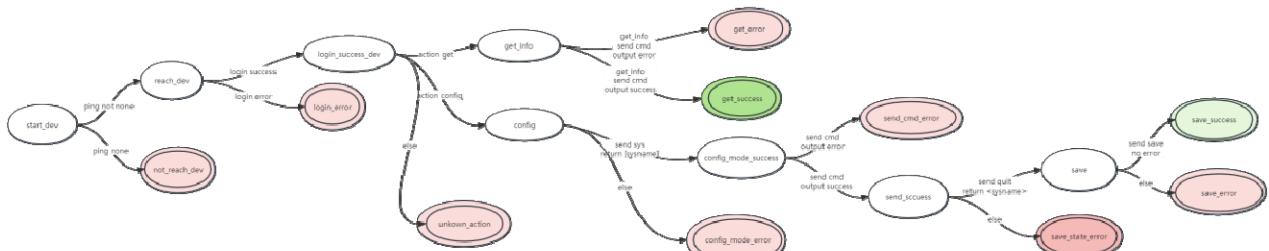


图4 获取网络信息配置有限状态机

通过自动化脚本或配置管理工具,对一系列设备执行远程登录,并批量下发预定义的配置命令^[10]。

5 应用与验证

上述研究成果已在搭建的试验环境中,进行了可行性验证。通过一系列测试案例,验证了在不同应用场景下的适用性和有效性。

在试验环境中,模拟了多样化的网络设备条件,包括不同型号的设备、多样的登录认证方式以及多种配置管理策略。通过应用研究成果,试验环境成功实现了对这些设备的集中统一管理。具体来说,研究成果支持了对不同型号网络设备的自动识

别和适配,无论设备来自哪个厂商或遵循何种标准,均能实现无缝接入和统一管理。

此外,研究成果还包括了对网络设备运维信息的高效采集,通过自动化的监控和日志分析,实时收集设备的运行状态、性能指标和故障信息。这些数据为网络的运营管理提供了信息支持,使得网络的维护和优化更加科学和高效。

6 结束语

现有网管系统在设备异构性、配置管理不一致性及设备批处理方面存在局限。针对这些问题,本文提出了一种基于FSM的网络设备信息采集技

术，能够实现异构网络设备的统一管理和配置，以适应网络设备数量激增，型号多样化，为网络管理带来了新的挑战^[11]。

本文提出的基于有限状态机的网络设备信息采集技术已被成功应用于实际网络环境。该技术能够有效地处理网络中存在的设备异构性问题，实现对数千台不同型号和类型的网络设备的集中式管理和配置。为大规模网络设备的管理与配置提供了一种高效、可靠的解决方案。

参考文献：

- [1] 王莉军, 隋欣. 基于 Python 实现的网络设备配置文件备份设计[J]. 中国新通信, 2023, 25(7): 31-32.
- [2] 张文川. 基于 SNMP 协议的计算机网络自诊断系统设计[J]. 自动化技术与应用, 2022, 41(6): 46-50.
- [3] 杨子, 王海港, 刘路登, 等. 基于有限状态机的电网

设备状态库研究[J]. 安徽电气工程职业技术学院学报, 2022, 27(1): 92-99.

- [4] 张莹. SNMP 在校园网络中的应用与研究[J]. 互联网周刊, 2022(5): 40-42.
- [5] 赖文书. 不同厂商网络设备配置备份脚本改进之路[J]. 网络安全和信息化, 2022(2): 145-148.
- [6] 孙晨阳. 网络设备配置信息自动备份系统的开发[J]. 中国水运, 2014(4): 38-39.
- [7] 解艳. 基于 SNMP 的网络信息管理系统设计[J]. 自动化技术与应用, 2021, 40(11): 94-97.
- [8] 胡文欣. 基于有限状态机的应用综述[J]. 电子元器件与信息技术, 2021, 5(10): 7-9.
- [9] 吴丹. 自动生成网络设备配置命令[J]. 网络安全和信息化, 2018(9): 52-56.
- [10] 杜巍. H3C 网络设备配置和维护[J]. 电子技术与软件工程, 2018(11): 20.
- [11] 陈军, 张伟刚, 黄立. 网络设备自动备份工具的实现[J]. 湖州师范学院学报, 2013, 35(3).

(上接第 129 页)

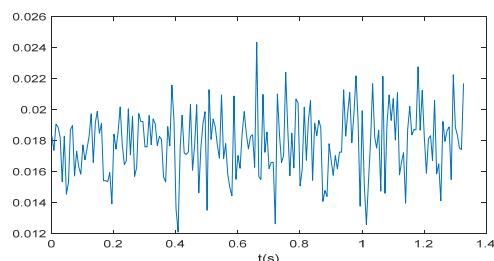


图 3 目标原始距离像相关系数曲线

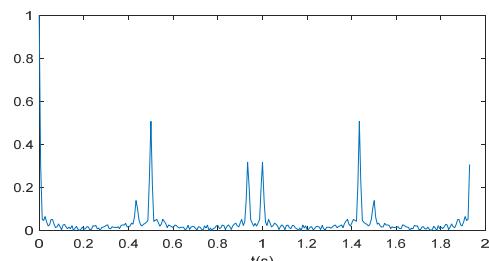


图 4 目标对齐距离像相关系数曲线

从上述仿真图形可以看出，雷达原始一维距离像由于目标自身运动等原因，散射点分布也随机变化。而且多个散射点散射强度相差不大，采用单个强散射点不能对齐雷达距离像回波，不能看出各个散射点的正弦运动曲线。采用散射点重心对齐雷达距离像，每次回波对齐有一个固定不变的散射点重心作为对齐的依据，因此散射点距离像可以对齐，3 个散射点各自形成 3 条正弦变化曲线，最后可以用距离像相关系数等方法估计目标的微动周期等参数。

6 结束语

空间飞行目标的微动特征参数是雷达跟踪测量目标时，雷达对空间目标进行类别识别的重要数据信息。对于宽带回波测量雷达而言，根据目标散射中心的后向散射回波信号，估算空间目标的周期参数是目标识别的一种重要手段。如何更好地抑制雷达测量空间目标回波信号中包含的随机误差的影响，提高雷达宽带信号估计空间目标运动周期精度，是值得进一步研究的课题。

参考文献：

- [1] 黄培康, 殷红成, 许小剑. 雷达目标特性[M]. 北京: 电子工业出版社, 2005.
- [2] 陈行勇, 陈海坚, 王林. 弹道导弹目标回波建模[J]. 现代雷达, 2010 (3): 27-31.
- [3] 张群, 罗迎, 何劲. 雷达目标微多普勒效应研究概述 [J]. 空军工程大学学报, 2011, 12(2): 22-24.
- [4] 高红卫, 谢良贵, 文树梁. 目标微动特性的微多普勒分析与仿真研究 [J]. 系统仿真学报, 2009, 21(4): 954-957.
- [5] 孙照强, 李宝柱, 鲁耀兵. 进动目标的微多普勒研究 [J]. 系统工程与电子技术, 2009, 31(3): 536-540.
- [6] 曹文杰, 张磊. 基于瞬时频率估计的进动锥体目标为多普勒频率提取方法[J]. 电子与信息学报, 2015, 37(5): 1091-1096.
- [7] 杨有春, 童宁宁. 弹道中段目标回波平动补偿与微多普勒提取方法[J]. 信息学报, 2013, 43(9): 1172-1182.
- [8] 牛杰, 刘永祥. 一种基于经验模型分解的锥体目标微动特征提取方法[J]. 电子学报, 2011, 39(7): 1712-1715.

多头注意力机制的卫星通信抗干扰波形决策方法

郑重^{1,2}, 管芸笛^{1,2*}, 李文吉^{1,2}, 郑寒雨^{1,2}, 乘珊^{1,2}, 蒲明龙^{1,2}, 金怡丰^{1,2}

1. 中国空间技术研究院 通信与导航卫星总体部, 北京 100094

2. 国家航天局卫星通信系统创新中心, 北京 100094

摘要:设计了基于波形决策的卫星通信抗干扰架构,根据干扰样式与参数,决策出收发双方应使用的最优波形,保证信息传输的可靠性和准确性。针对干扰样式与最优传输波形映射关系复杂的问题,有效挖掘干扰特征与波形抗干扰能力,提出了一种基于多头注意力机制的抗干扰波形决策算法,主要包含特征构建、特征增强和匹配决策三个模块,设计多头自注意力机制和多层次感知机神经网络实现特征增强与匹配决策,并对多头的数量进行了优化。仿真结果表明基于多头注意力机制的抗干扰波形决策算法经足够轮次的训练后,能达到稳定收敛状态,且决策准确率能够达到99%,实现卫星通信系统通过波形的切换有效应对干扰。

关键词:卫星通信;抗干扰;波形决策;多头注意力机制

文献标志码: A

中图分类号: TP391

Waveform Decision-Making Method Based on Multi-head Attention Mechanism for Anti-jamming Satellite Communication

ZHENG Zhong^{1,2}, GUAN Yundi^{1,2*}, LI Wenji^{1,2}, ZHENG Hanyu^{1,2}, LUAN Shan^{1,2}, PU Minglong^{1,2}, JIN Yifeng^{1,2}

1. Institute of Telecommunication and Navigation Satellites, China Academy of Space Technology, Beijing 100094, China

2. Innovation Center of Satellite Communication System, China National Space Administration, Beijing 100094, China

Abstract: A waveform decision-making anti-jamming architecture for satellite communication is designed to decide the optimal waveforms that should be used by both the transmitter and receiver according to the interference pattern and parameters, so as to ensure the reliability and accuracy of information transmission. Aiming at the problem of complex mapping relationship between jamming style and optimal transmission waveform, and effective mining of jamming features and waveform anti-jamming ability, an anti-jamming waveform decision-making algorithm based on multi-head attention mechanism is proposed, which mainly contains three modules, feature construction, feature enhancement and matching decision. Multi-head attention and multi-layer perceptron neural network are designed to realize feature enhancement and matching decision, and the number of multi-heads is optimized. The simulation results show that the anti-jamming waveform decision-making algorithm based on multi-head attention mechanism can reach a stable convergence state after enough rounds of training, and the decision-making accuracy can reach 99%, which realizes the satellite communication system to effectively deal with the jamming through the switching of waveforms.

Key words: satellite communication; anti-jamming; waveform decision; multi-head attention

随着星上处理能力的发展,为有效应对干扰,通信卫星从传统的透明转发模式逐渐升级至星上

作者简介:郑重(1990—),男,博士,高级工程师,研究方向为卫星通信与网络;管芸笛(1997—),通信作者,女,硕士,工程师,研究方向为卫星通信与网络、人工智能抗干扰,E-mail:guanyd1002@126.com。

再生处理模式。与此同时，针对越来越复杂多变的传输环境与多样化的传输需求，卫星可支持多种体制的传输波形。在实际应用中，需要根据复杂电磁环境的特性和终端的业务需求，决策出最优的通信波形，从而更加合理地利用传输资源，并保证信息传输的可靠性和准确性。

本文针对抗干扰场景，设计了基于波形决策的卫星通信抗干扰架构，根据干扰样式与参数，决策出收发双方应使用的最优波形，保证信息传输的可靠性和准确性。针对干扰样式与最优传输波形映射关系复杂的问题，有效挖掘干扰特征与波形抗干扰能力，提出抗干扰波形决策算法，并进行仿真验证。

1 系统模型和问题描述

1.1 基于波形决策的抗干扰卫星通信系统

针对卫星与用户终端之间的通信，上行干扰仍是实施最为简单、效果最为显著的干扰手段。地基上行干扰的干扰样式包括单音、多音、部分频带、全频带、扫频、跟踪转发式干扰等。有效应对干扰的方式是通过切换波形样式与参数，针对特定干扰收发双方使用最有效的波形进行传输，实现最佳的抗干扰效果。通过卫星通信系统上配置的电磁环境感知设备，感知当前通信传输链路的电磁频谱环境，分析当前的电磁频谱态势，提取电磁环境参数，并识别电磁干扰样式，最终将干扰样式、干扰功率、干扰频点等干扰参数信息传递至波形决策系统。

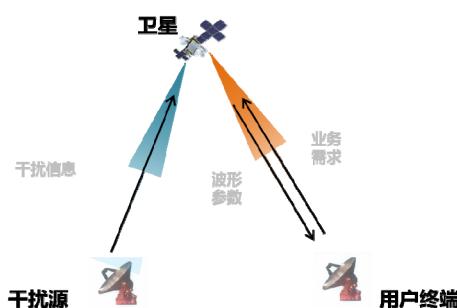


图1 干扰场景下卫星与用户终端传输示意图

卫星通信系统可配置多种体制的传输波形，其波形参数包括但并不限于调制方式、扩频方式、分集方式、信道编码方式、交织、加扰等。为支撑后续波形决策，需构建特定电磁环境与通信波形参数

相关联的知识库，将该知识库作为后续波形决策的先验信息。该知识库需要根据特定的电磁环境、特定的通信波形参数，计算出对应的误码率。将电磁环境与通信波形参数通过误码率指标相关联。

用户方根据所需的通信业务需求向波形决策系统提交业务类型与传输速率需求。业务需求包括但并不限于多媒体业务、数据业务、语音业务等。传输速率需求为满足该通信业务所需的最低传输速率值。

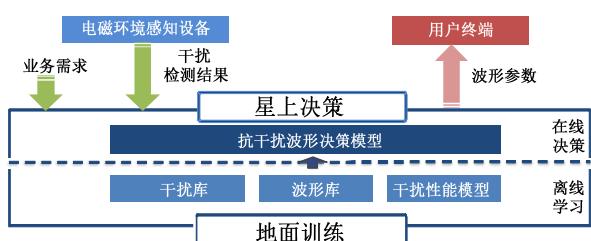


图2 卫星通信抗干扰波形决策系统架构

波形管理系统将波形参数通过单向广播的方式传输终端，终端更新波形。波形广播帧结构如图3所示，由前导序列、广播时隙、信令时隙和业务时隙组成。前导序列用于实现定时恢复和载波恢复，I/Q路均可采用全0符号或walsh码。广播时隙用于下行广播信息的传输，采用低阶的编码调制方式，保证波束内所有终端可靠地接收到广播信息。广播时隙的数量需要根据波束的个数及广播信息量进行优化设计。信令时隙用于指示下行物理帧数据的组成方式，包括数据的载波速率、调制编码方式、扩频比、起始和结束时隙序号，以及数据长度等信息。采用低阶的编码调制方式，保证波束内所有终端可靠地接收到信令信息。每个业务时隙包括导频和数据体。导频用于高阶调制时辅助载波恢复，以均匀方式插入数据体中，每N个数据符号插入M个导频符号，导频为I/Q路全零的数据。

最后，卫星接收机通过新的波形参数进行接收，成功接收后，系统切换至新的波形进行工作。

1.2 问题描述

基于以上系统，本文建立如图4所示的抗干扰决策通信模型。

原始信号经发送端进行编码、交织、跳频、加扰、调制等技术，形成传输波形信号。发送端不同

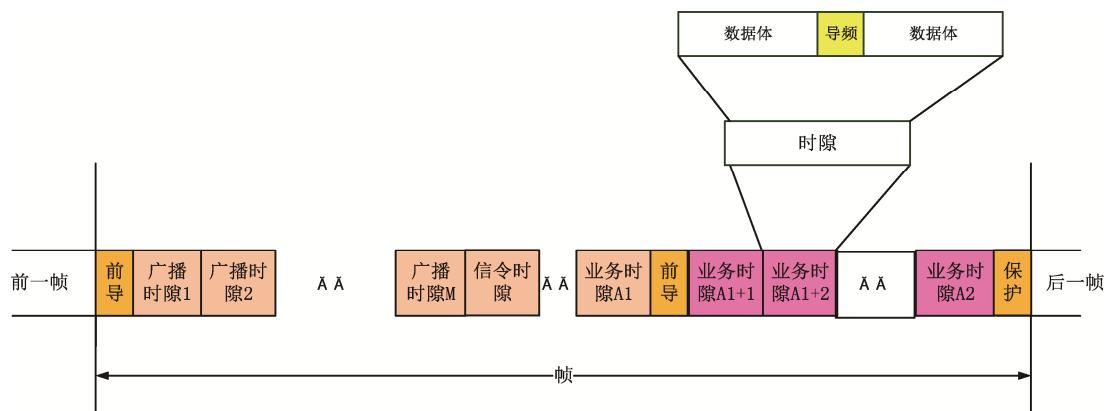


图 3 波形广播帧结构示例

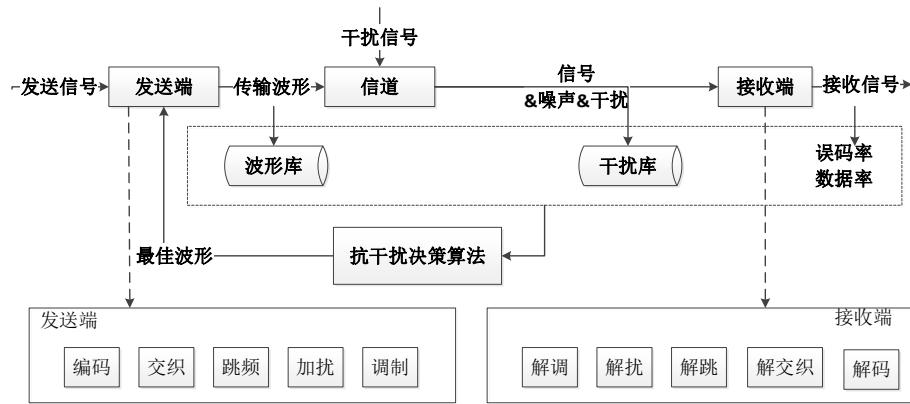


图 4 抗干扰决策通信模型

抗干扰技术组成波形 $w = \{Cod, Mod, \dots\}$ ，不同波形构成了波形库 $W = \{w_1, w_2, \dots, w_n\}$ 。传输过程中，不同样式和强度的干扰信号与信道中噪声共同作用于传输波形。信号接收前需要进行干扰检测识别，以提取本次传输中的干扰特征，如干扰类型、干信比、干扰带宽等，构成一种干扰 $j = \{J_{type}, JSR, \dots\}$ ，不同干扰构成干扰知识库 $J = \{j_1, j_2, \dots, j_m\}$ 。接收端对信号进行对应的解调、解扰、解跳、解交织、解码等操作恢复信号，统计错误比特个数并计算传输误比特率 BER 。本文将一次传输受到的干扰 j_i 与信噪比 SNR 作为传输环境，波形在此传输环境下的误比特率 BER 作为传输性能的衡量指标 $r = \{SNR, BER\}$ 。

因此，在一次通信过程中，就可以建立 r_i 、 j_i 与 w_k 之间的匹配关系，这就意味着在干扰 j_i 下，想要达到 r_i 这样的传输性能，采用的最佳波形是 w_k 。

为最大化信号传输性能，减少干扰带来的影响，考虑根据干扰情况 j_i 以及用户的性能需求 r_i ，

通过卫星通信抗干扰波形决策算法进行抗干扰波形的自主决策，即选择出一种速率的波形，能够抵抗当前类型和强度的干扰，使传输波形达到需求的性能。

2 基于多头注意力机制的波形决策算法

本章研究基于多头注意力机制的卫星通信抗干扰波形决策算法，其训练样本由波形库、干扰库和抗干扰性能库中的数据构成。算法整体架构由图 5 给出。本文的核心任务可定义为：针对输入的干扰环境和此干扰下传输需求的性能 $u_i = (j_i, r_i)$ ，选择出一种最佳波形 w_k ，能够最大程度地满足当前传输性能需求。

算法主要包含特征构建、特征增强和匹配决策三个模块。特征构建模块需要将初始波形、干扰和性能等不同表达形式的元素映射为向量形式 w_k 、 j_i 和 r_i ，以便神经网络的进一步处理。特征增强部分将学习中更为有效的特征，并赋予其更高权重，以提升匹配准确率，得到增强向量 w'_k 、 u'_i 。匹配

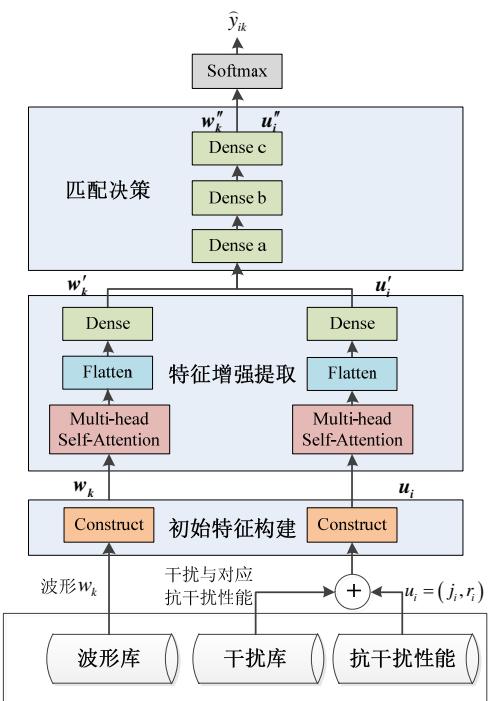


图 5 基于多头注意力机制的抗干扰波形决策网络

决策模块将进一步凝练特征得到 w_k'' 、 u_i'' ，并计算 w_k'' 与 u_i'' 的匹配得分 \hat{y}_{ik} ， \hat{y}_{ik} 越高则对应的波形越匹配当前干扰，即应选择最高 \hat{y}_{ik} 对应的波形 w_k 。三个模块计算可表示为：

$$\begin{aligned} U_1, W_1 &= f^1(U, W) \\ U_2, W_2 &= f^2(U_1, W_1) \\ Y &= f^3(U_2, W_2) \end{aligned} \quad (1)$$

2.1 初始特征构建

波形中包含调制编码方式、干扰中包含干扰类型等非向量特征，无法通过神经网络进行处理。因此需要对波形、干扰的初始特征进行构建。设计类别编码进行非向量特征的映射，对数值特征进行归一化，并拼接所有特征，形成向量形式的特征表示。波形构建过程如图 6 所示，干扰特征构建与之类似，主要包含对干扰类型、干信比、干扰带宽/音数等参数。

2.2 特征增强提取

初始构建的干扰性能与波形特征包含内容信息较简单，表征能力较弱，因此，需要对其进行增强提取，进一步挖掘波形、干扰特征关系，并增强其有用特征的权重，实现针对性的匹配决策。

自注意力机制常用于远距离特征的关系挖掘和重要特征的自动加权，目前已广泛应用于自然语言处理、图像、音频、空间等各方向的特征提取。因此，本研究采用多头自注意力机制，分别对波形 w_k 与干扰性能 u_i 的初始特征进一步提取增强。以波形向量 w_k 为例， $w_k = [w_{k,1}, w_{k,2}, \dots, w_{k,n}]^T$ 包含 n 个特征，赋予每一个特征 $w_{k,j}$ 三个向量，通过输入 w_k 分别乘以权值矩阵 W_Q 、 W_K 、 W_V ，再加上对应偏置得到。

$$\begin{aligned} K &= W_K X + b_K \\ V &= W_V X + b_V \\ Q &= W_Q X + b_Q \end{aligned} \quad (2)$$

对每一个特征计算相似性，通过点积拟合得到 $\text{Similarity}(Q, K)$ ，经过归一化后再与 Value 相乘得到每个输入特征向量的评分。计算方式为：

$$\text{Similarity}(Q, K) = Q \times K^T \quad (3)$$

最终得到 Attention 评分为：

$$\text{Attention}(Q, K, V) = \text{Similarity}(Q, K)V \quad (4)$$

本研究中引入多头注意力机制，即将自注意力机制的计算过程做 h 次，每次即一个“头”。其中每次参数 Q 、 K 、 V 进行变换时的参数 W_Q 、 W_K 、 W_V 不同，从而允许模型在不同的表示子空间里学到相关特征。拼接 h 次 Attention 评分值，最后再进行一次线性变换得到多头注意力的结果。每一头的计算方式即：

$$h_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V) \quad (5)$$

多头注意力的结果即：

$$\text{multiHead}(Q, K, V) = \text{Concat}(h_1, h_2, \dots, h_h) \quad (6)$$

2.3 匹配决策网络

匹配决策目的是计算出输入的干扰需求与波形的匹配得分，从而在波形库选出最适配的一个，即既能满足当前抗干扰需求，又不会造成资源浪

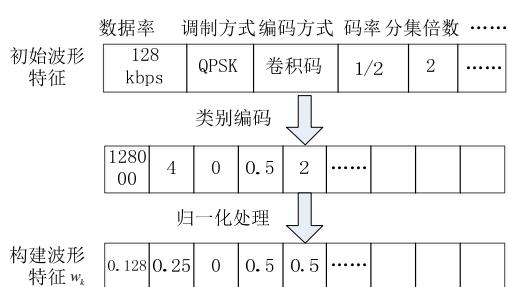


图 6 波形特征构建示意图

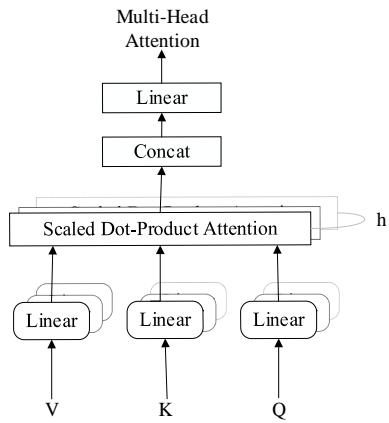


图 7 多头注意力机制模型

费。因此，采用多层感知机神经网络构建匹配决策网络模块，拟合波形与干扰和需求的匹配程度。

多层感知机由输入层、 l 层全连接层和输出层构成。本文采用3层激活函数为LeakyReLU的全连接层。输入为 w_k'' 、 u_i'' 的串联向量，输出Softmax归一化后的匹配得分 \hat{y}_{ik} 。对于每个 u_i'' ，计算它与所有 w_k'' 的 \hat{y}_{ik} ，再进行排序，最高 \hat{y}_{ik} 对应的波形即最优决策结果。

2.4 优化目标设计

在训练过程中，对于每个 u_i ，其最佳波形为 w_{pos} ，则 (u, w_{pos}) 为一个正样本，那么设置其余波形中任一个为负样本 (u, w_{neg}) 。模型计算出正样本 (u, w_{pos}) 与负样本 (u, w_{neg}) 经过抗干扰波形决策网络的匹配概率 \hat{y}_{pos} 与 \hat{y}_{neg} ，并计算如下交叉熵损失：

$$L_{\text{CE}} = -\frac{1}{N_s} \sum_{n=1}^{N_s} \left(\ln(\hat{y}_n^{\text{pos}}) + \ln(1 - \hat{y}_n^{\text{neg}}) \right) \quad (7)$$

其中， N_s 为一次训练中的总样本数。

在整个训练过程中，通过最小化损失函数来不断梯度更新神经网络模型的参数。在每轮训练后的测试中，将此时的模型运用于测试集，即计算模型在未接触过的样本集上的决策结果，并计算决策准确率：

$$acc = \frac{N_{\text{right}}}{N_{\text{Sample}}} \quad (8)$$

将决策准确率作为模型性能衡量指标，其中 N_{Sample} 为测试集的样本数量， N_{right} 为决策正确的样本数量，当一次预测中预测的最佳波形与样本标签波形一致时，即视为决策正确。

3 仿真结果与分析

本章对提出的算法决策性能进行仿真。首先仿真不同超参数对决策性能的影响，然后对比本文所提算法与随机策略、协同滤波算法的决策性能。

3.1 参数设置

本实验设计共21种波形，干扰需求共40 800种。具体参数取值由表1给出。

表 1 波形、干扰与需求参数设置

	参数类型	参数取值
干扰	干扰类型	部分带、全频带
	干扰带宽占比	0:0.1:1
	干信比/dB	0:5:60
	调制方式	QPSK、BPSK
	需求	信道
		误码率
波形	需求	AWGN
	波形	编码方式
		卷积、LDPC
		码率
需求	波形	分集重数
		传输速率/kb/s

决策网络与仿真参数具体设置如表2所示。

表 2 网络与仿真参数设置

	参数类型	参数取值
学习率	训练：测试	0.001
注意力机制头数	训练轮次	8:2
1、3、5、7	1 000	

3.2 结果分析

为验证本文所提决策算法的有效性，本实验基于Python3.7，利用深度学习框架Tensorflow2.5构建网络模型，训练采用Adam优化器，训练批次数batch_size=128。为体现本文所提算法的决策性能优势，与其他类似决策方法进行了仿真对比，包括随机决策和基于协同滤波的决策。算法的决策准确率变化曲线如图8所示，基于多头注意力机制的抗干扰波形决策算法经足够轮次的训练后，能达到稳定收敛状态，且决策准确率能够达到99%，明显高于基于协同滤波和随机决策。由此可得，采用多头注意力机制能够有效强化特征，有效提升决策性能。

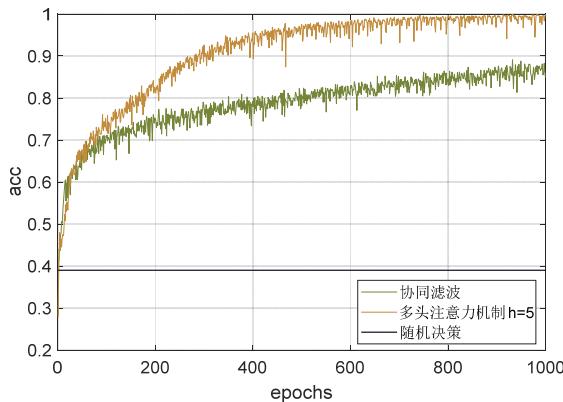


图 8 不同决策算法性能对比

为得到最佳决策模型,本文对不同注意力机制的多头数 h 对决策性能的影响进行了对比,如图 9 所示。可以看出 h 越大决策的准确率越高。但随着 h 增大,注意力变换次数增加,算法复杂度增大,会造成决策速度下降。当 $h=7$ 时,对算法性能的增益可以忽略。当 $h=5$ 时,已能够实现 99.82% 的决策准确率。

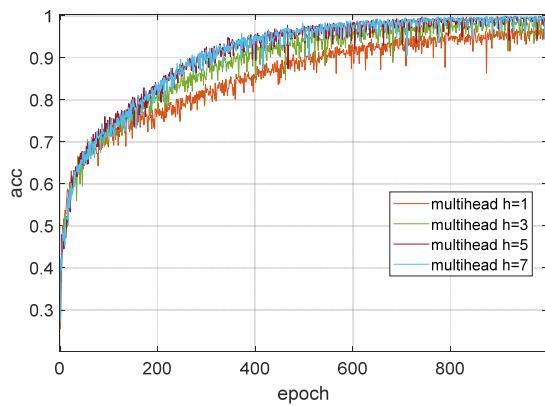


图 9 不同头数对决策性能的影响

4 结束语

针对干扰环境下卫星通信系统的最优波形决策问题,本文首先设计了基于波形决策的卫星通信抗干扰架构,针对干扰样式与最优传输波形映射关系复杂的问题,提出了基于多头注意力机制的波形决策算法。该算法将干扰环境与传输需求和通信波形的匹配关系建模为决策网络的输入输出,并引入多头自注意力机制进行特征的强化提取,以学习波形与干扰间的隐藏特征。最后通过仿真对比,验证了本文所提算法在决策性能上的优越性,并优化了最佳注意力机制的多头数,均衡了算法性能

与计算开销。在后续工作中,将进一步研究在未知复杂干扰下波形的决策问题,不断强化卫星通信抗干扰能力。

参考文献:

- [1] 姚富强. 通信抗干扰工程与实践[M]. 北京: 电子工业出版社, 2012.
- [2] WANG H, CHANG Q, XU Y. An integrated beam anti-jamming algorithm for low-Orbit navigation augmentation[J]. IEEE Communications Letters, 2022, 26(4): 877-881.
- [3] HE K, CHENG Y, DUAN H, et al. Research on strategy-based anti-jamming intelligent decision engine for adaptive burst communication[C]//2020 IEEE 20th International Conference on Communication Technology. IEEE, 2020.
- [4] YAO B, WANG, HOU D, et al. Anti-follower jamming wide gap multi-pattern frequency hopping communication method[J]. Defence Technology, 2020, 16(2): 453-459.
- [5] LIU H, LI G. OFDM-based broadband wireless networks: design and optimization[M]. New York: John Wiley & Sons Ltd., 2005.
- [6] ZHOU Q, LI Y, NIU Y. Intelligent anti-Jamming communication for wireless sensor networks: a multi-agent reinforcement learning approach[J]. IEEE Open Journal of the Communications Society, 2021, 2(99): 775-784.
- [7] JIANG W, REN Z, HUANG K, et al. A joint space-frequency anti-jamming scheme based on reconfigurable intelligent surface[C]//2022 IEEE Global Communications Conference (GLOBECOM 2022), Rio de Janeiro, Brazil, 2022: 6408-6414.
- [8] 李少谦, 程郁凡, 董彬虹, 等. 智能抗干扰通信技术研究[J]. 无线电通信技术, 2012, 38(1): 1-4.
- [9] 刘猛, 戚红雨, 王荆宁, 等. 基于神经网络算法的智能抗干扰系统设计[J]. 计算机测量与控制, 2018, 26(10): 155-159.
- [10] YE F, ZHOU Z, TIAN H, et al. Intelligent anti-jamming decision method based on the mutation search artificial bee colony algorithm for wireless systems[C]//2019 USNC-URSI Radio Science Meeting, IEEE, (Joint with AP-S Symposium), 2019: 27-28.
- [11] 冉雨, 陈大勇, 程郁凡, 等. 基于改进人工蜂群算法的认知抗干扰智能决策技术研究[J]. 信号处理, 2019, 35(2): 240-249.
- [12] 桑万超, 高晖. 面向无人机网络的通信感知一体化的高效能波形选择方法[J]. 无线电通信技术, 2023,

- 49(1): 133-142.
- [13] PU Z, NIU Y, ZHANG G. A multi-parameter intelligent communication anti-jamming method based on three-dimensional q-learning[C]//2022 IEEE 2nd International Conference on Computer Communication and Artificial Intelligence (CCAI), Beijing, China, 2022: 205-210.
- [14] ZHANG L, MA L, TIAN F, et al. An anti-jamming intelligent decision-making method for multi-user communication based on deep reinforcement learning[C]//2022 IEEE 22nd International Conference on Communication Technology (ICCT), Nanjing, China, 2022: 1335-1339.
- [15] 吴尚峰, 唐麒, 魏急波, 等. 具有吞吐量和时延约束的软件无线电波形部署决策算法[J]. 国防科技大学学报, 2015, 37(1): 136-141.
- [16] LI X, FANG J, CHENG W, et al. Intelligent power control for spectrum sharing in cognitive radios: a deep reinforcement learning approach[J]. IEEE Access, 2018, 6: 25463-25473.
- [17] ZHU Q, QU Z, QIAO J, et al. Decision-making method for communication parameter selection via support vector regression[C]//2018 Eighth International Conference on Instrumentation & Measurement, Computer, Communication and Control (IMCCC), Harbin, China, 2018: 370-374.
- [18] KE C, LI J, CHENG W, et al. An intelligent anti-interference communication method based on game learning[C]//2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), Chongqing, China, 2021: 182-186.

(上接第 204 页)

跃变和平均相位梯度最小,最大程度满足了相位连续性假设。通过仿真实验进一步验证了方法的有效性,与传统 Delaunay 三角网及基于时间相干性更新的APSP 网络相比,提高了相位解缠的精度,为之后解缠算法的改进提供了新思路。

参考文献:

- [1] 朱建军, 李志伟, 胡俊. InSAR 变形监测方法与研究进展[J]. 测绘学报, 2017, 46(10): 1717-1733.
- [2] 王彦平, 白泽朝, 林赟, 等. InSAR 双向矩形角反射器阵列形变监测精度评估与验证[J]. 武汉大学学报(信息科学版), 2021, 46(10): 1471-1477.
- [3] 白泽朝, 吕森, 曹琨, 等. 结合星载地基 InSAR 分析的四川新磨村滑坡早期识别与灾后监测[J]. 地球物理学进展, 2020, 35(5): 1637-1643.
- [4] SHI X, JIANG H, ZHANG L, et al. Landslide displacement monitoring with split-bandwidth interferometry: a case study of the shuping landslide in the Three Gorges area[J]. Remote Sensing, 2017, 9(9): 937.
- [5] LUO H, LI Z, CHEN J, et al. Integration of range split spectrum interferometry and conventional InSAR to monitor large gradient surface displacements[J]. International Journal of Applied Earth Observation and Geoinformation, 2019, 74: 130-137.
- [6] YU H, LAN Y, YUAN Z, et al. Phase unwrapping in InSAR: a review[J]. IEEE Geoscience and Remote Sensing Magazine, 2019, 7(1): 40-58.
- [7] ITOH K. Analysis of the phase unwrapping problem[J]. Applied Optics, 1982, 21: 2470.
- [8] GHIGLIA D C, ROMERO L A. Minimum LP-norm two-dimensional phase unwrapping[J]. Journal of the Optical Society of America A, 1996, 13(10): 1999-2013.
- [9] PRITT M D, SHIPMAN J S. Least-squares two-dimensional phase unwrapping using FFT's[J]. IEEE Transactions on Geoscience & Remote Sensing, 1994, 32(3): 706-708.
- [10] COSTANTINI M, ROSEN P A. A generalized phase unwrapping approach for sparse data[C]//IEEE 1999 International Geoscience and Remote Sensing Symposium. IGARSS'99, 1999: 267-269.
- [11] 马张烽, 蒋弥, 李桂华, 等. 空间网络对时序 InSAR 相位解缠的影响——以 Delaunay 与 Dijkstra 网络为例[J]. 测绘学报, 2022, 51(2): 248-257.
- [12] ZHANG L, DING X, LU Z. Modeling PSInSAR time series without phase unwrapping[J]. IEEE Transactions on Geoscience Remote Sensing, 2010, 49(1): 547-556.
- [13] SONG X, ZHANG L, LU Z, et al. Towards retrieving discontinuous deformation of bridges by MTInSAR with adaptive segmentation[J]. IEEE Transactions on Geoscience Remote Sensing, 2024.
- [14] LUO H, LI Z, DONG Z, et al. A new baseline linear combination algorithm for generating urban digital elevation models with multitemporal InSAR observations[J]. IEEE Transactions on Geoscience Remote Sensing, 2019, 58(2): 1120-1133.
- [15] MA Z F, JIANG M, Khoshmanesh M, et al. Time series phase unwrapping based on graph theory and compressed sensing[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 1-12.
- [16] SHANKER A P, ZEBKER H. Edgelist phase unwrapping algorithm for time series InSAR analysis[J]. Journal of the Optical Society of America A, Optics, Image Science & Vision, 2010, 27(3): 605-612.

一种基于插件模式的网络自动化系统设计与实现

熊宗炬，李朝阳，李敬歌，张万彬，张小锋

中国西昌卫星发射中心，四川 西昌 615000

摘要：传统自动化运维工具在应对日益复杂的网络管理问题时面临诸多困难。因此，提出一种插件化、高并发、高可靠的网络自动化平台架构方法，并在 python 生态下编程实现为一款自动化运维软件。通过对该软件在仿真网络中相关性能数据的对比分析，表明基于该方法的自动化运维软件在灵活性、可扩展性和执行效率等方面相较传统工具优势明显，具有良好的应用前景和推广价值。

关键词：网络自动化；插件系统；SSH；Paramiko；Scrapli

文献标志码：A

中图分类号：TP311.11

Design and Implementation of Network Automation System Based on Plugin Mode

XIONG Zongju, LI Zhaoyang, LI Jingge, ZHANG Wanbin, ZHANG Xiaofeng

Xichang Satellite Launch Center, Xichang, Sichuan 615000, China

Abstract: Traditional automated operation and maintenance tools face many difficulties in dealing with increasingly complex network management problems. Therefore, this paper proposes a plug-in, high-concurrency, and high-reliability network automation platform architecture method, and implements it as an automated operation and maintenance software in the python ecosystem. Through the comparative analysis of the relevant performance data of the software in the simulated network, it shows that the automatic operation and maintenance software based on this method has obvious advantages over traditional tools in terms of flexibility, scalability and execution efficiency, and has good application prospects and promotion value.

Key words: network automation; plug-in system; SSH; Paramiko; Scrapli

自计算机网络诞生之日起，网络便是复杂且难以管理的。这种复杂性主要体现在网络设备的多样性、网络管理方式的不同以及生产网络中多种网管系统并存等方面。为应对上述因素导致的复杂网管现状，必须寻求一种更加开放、灵活和高效的管理方式，构建以网络开放可编程与自动化技术为基础的智简网络。

当前，运维自动化领域的诸多平台或工具（譬如 Ansible、Salt）^[1]的功能正逐步向网络自动化领域扩展。但是，这些平台起初的管理对象和任务主要是对服务器及构建于其上的应用程序的配置管理，它们在对网络设备进行管理方面往往存在任务脚本调试困难、管理国内厂商设备缺乏插件以及学习曲线陡峭等诸多缺陷^[2]。因此，本文提出一种基于插件开发运行模式的、高并发、高可靠的网络自

动化平台架构方法，以应对生产网络中日益复杂的网络管理问题。

1 网络自动化架构综述

网络自动化并不是简单地将任务置于一个循环中进行计算和运行，而是一个非常复杂的问题。它是分布式的、高并发的、实时的、高可用的、性能敏感的、安全敏感的复杂控制系统，它需要一个全方位的系统架构^[3]。

下面从设备管理方式、软件部署方式、信息传输方式、数据格式约定、软件功能扩展以及任务执行模式等六个层面对运维自动化系统架构作一般性阐述。（1）设备管理方式包括基于代理（agent-based）和无代理（agentless）两种。前者适合围绕服务器开展的各类自动化运维，后者适合网络设备的

自动化运维。(2) 软件部署方式包括集中化和去集中化,通常采取基于代理方式的自动化工具是集中化部署,而无代理方式则对应去集中化。(3) 信息传输方式包括私有定制协议和标准协议两种,在基于代理方式的自动化工具中多采用私有定制协议实现代理端与服务端之间的信息交互。虽然多数网络操作系统不支持代理方式,但几乎都支持标准的SSH协议。(4) 数据格式约定包括DSL(domain-specific language)和通用数据格式标记语言两种。与专门的DSL相比较,大多数自动化工具采用YAML作为数据格式标记语言。(5) 软件功能扩展是指自动化工具一般采取插件化运行模式,通过高级脚本语言实现软件功能的自由扩展,这一点对于国产网络设备的管理尤为重要。通常的高级脚本语言包括Ruby和python。(6) 任务执行模式包括push、pull以及事件驱动三种模式。push模式是指配置管理信息从同一个地方被推送到被管设备;pull模式是指配置管理信息从一个地方被下拉到被管设备;事件驱动模式是指特定事件触发后则会引起相应的配置管理操作^[4]。

本文在对传统自动化工具在网络自动化运维方面的主要特性进行归纳提炼的基础上(见表1),将进一步设计出支持幂等性^[8]、支持SSH、支持Jinja2、支持YAML编排、支持插件扩展、采用无代理方式管理设备和集中化软件部署方式的高并发网络自动化管理平台。

2 插件系统设计

本文设计面向网络自动化的高并发插件系统

表1 网络自动化运维平台架构分析

类别	Ansible ^[5]	Salt ^[6]	Stackstorm ^[7]
管理方式	agentless	agent-based/agentless	agent-based
传输协议	SSH	SSH/message-bus/NAPALM	message-bus
功能扩展	plugin-based	plugin-based	plugin-based
模板支持	Jinja2	Jinja2	不支持
幂等性	支持	支持	不支持

体系结构如图1所示。从功能组成层面而言,整个高并发插件系统包括基础类库、插件子系统、核心调度器、用户自定义模块及执行结果五部分。核心调度器是整个插件系统的关键核心部件,其内部逻辑决定着作业函数在一组网络设备上的执行流程,负责相关类库加载、插件注册、任务执行及结果重塑等。

基础类库是对网络自动化操作过程中的共性功能的抽象,主要包括配置解析模块、作业处理模块、过滤处理模块、异常处理模块和结果处理模块。其中,配置解析模块负责加载和解析各类用户配置文件;作业处理模块负责对用户自定义的作业函数进行抽象封装;过滤处理模块基于网络设备抽象类的属性或特定过滤函数对网络设备进行过滤,通常将一个用户自定义作业函数作用于过滤后的一组网络设备以完成某项自动化配置任务;异常处理模块定义整个插件系统所需异常类;结果处理模块将一个作业函数在一个主机上执行以及在一组主机

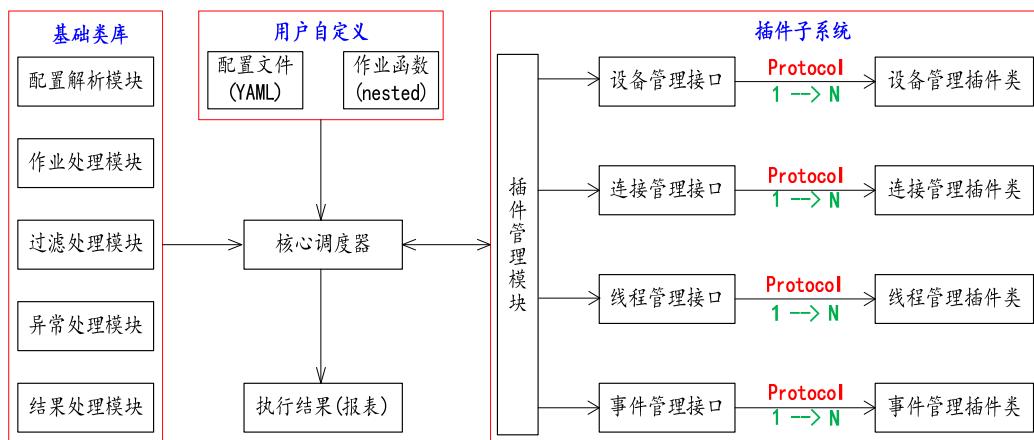


图1 系统体系结构设计

上执行的返回结果进行规范化处理。

插件子系统是实现整个软件平台功能可扩展性的基础支撑。其设计思想简述如下：在对网络设备自动化管理流程中的关键环节进行提炼分析的基础上，以继承 Protocol^[9]类的方式重点实现设备管理接口、连接管理接口、线程管理接口和事件管理接口。以上述各类接口类为基准，一方面对内通过插件管理模块为核心调度模块提供设备操作支持；另一方面对外为网管用户通过不同协议连接到多厂商设备执行不同管理配置任务而提供技术途径及标准。

以连接管理为例简要阐述本设计在软件功能可扩展方面的优势。首先，在连接管理插件接口定义过程中抽象出当前连接网络设备的主流协议类型。其次，连接管理插件类是对连接管理接口的具体实现，用户可根据实际需求实现多个连接类，譬如基于 Paramiko、Netmiko、Netconf 等不同类库实现面向不同厂商不同自动化级别的连接管理插件

类。最后，插件管理模块可通过用户在 yaml 配置文件中的设置实现不同连接管理插件程序的自动注册和自动加载，进一步提高了整个软件系统的可用度和灵活性。

3 插件系统实现

3.1 实现流程

面向网络自动化运维的高并发插件系统实现流程见图 2，主要包括解析配置文件、整合设备资源、自动化批处理和重塑结果信息四个阶段。

第一，解析配置文件阶段包括解析插件配置文件、解析设备配置文件和解析其他配置文件三个步骤。通过对插件配置选型的解析，使得核心调度器掌握当前所使用的设备管理插件类及其运行参数、连接管理插件类及其运行参数、线程管理插件类及其参数和事件管理插件类及其运行参数。通过对设备配置选项的解析，使得核心调度器掌握整个自动化运维系统运行的基础资产资源。通过对其他配置

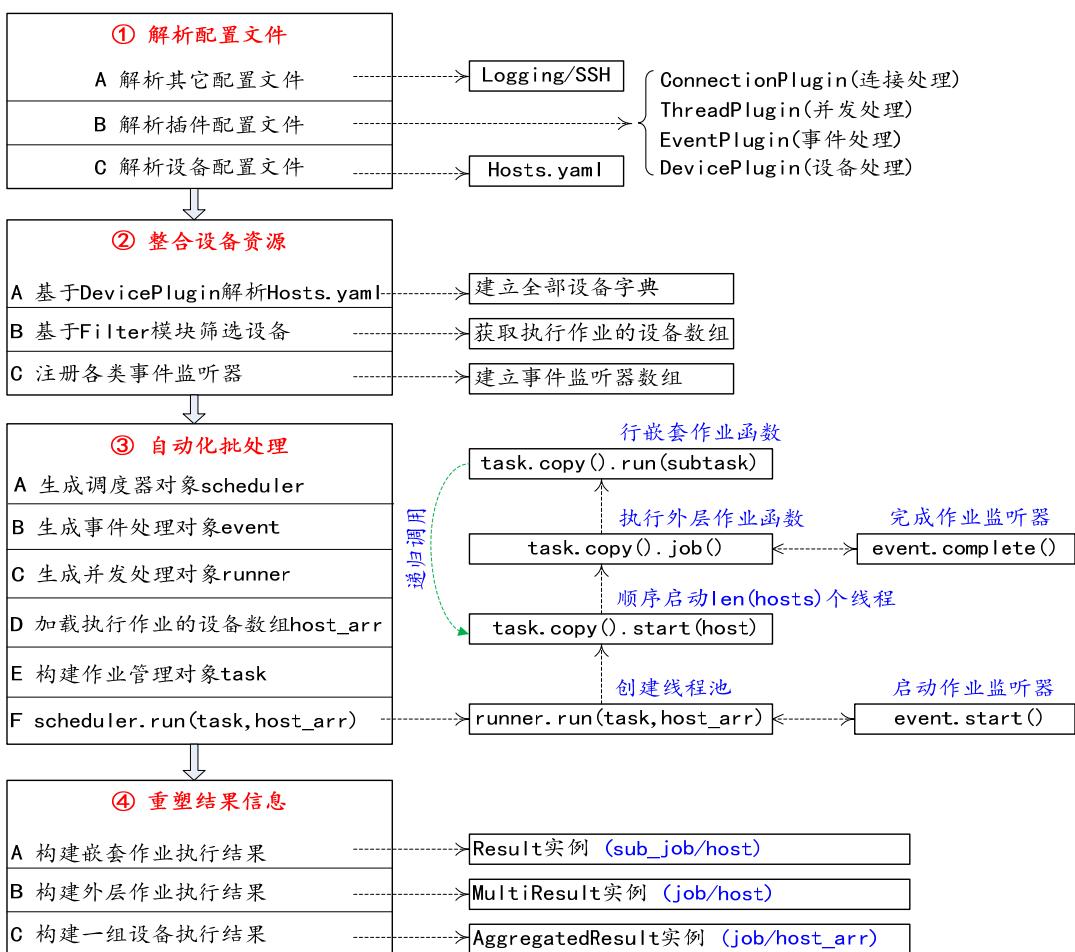


图 2 系统实现流程

选项的解析使得核心调度器获得诸如 SSH 协议配置及日志参数配置等其他设置。

第二，整合设备资源阶段包括三个主要步骤：

- (1) 基于当前加载的设备管理插件程序进一步解析获得将要执行特定作业函数的一组网络设备；
- (2) 借助过滤模块并以前期刻画的网络主机对象(Host)的属性和特定过滤函数为基础进一步对网络设备进行分类组合，为特定作业函数在不同设备分组上的批量执行打下基础；(3) 将各类事件监听器组合成一个数组并依附于特定的任务执行，以此实现整个系统基于事件驱动的操作管理。通常这类事件监听器主要关注特定作业执行前和执行后需要对网络设备完成的具体操作。

第三，自动化批处理阶段主要包括六个关键步骤，分别是生成核心调度器对象 scheduler、生成事件处理对象 event、生成并发处理对象 runner、加载需执行作业的网络设备所构成的数组 host_arr、构建作业管理对象 task、通过核心调度器在特定设备数组 host_arr 上运行某个特定网络运维作业。自动化批处理阶段是整个高并发插件系统的核心，文章在 3.2 节进一步阐述。

第四，重塑结果信息阶段主要包括构建嵌套作业执行结果、构建外层作业执行结果和构建一组设

备执行结果三个步骤。本文所设计的网络自动化运维系统为用户提供编写和执行作业函数所需接口及标准。作业函数分为不嵌套和嵌套两类。对于嵌套的内层作业函数将其执行结果存入 Result 实例，不嵌套的作业函数或嵌套的外层作业函数的执行结果存入由 Result 实例构成的数组中，特定作业函数在一组设备上的执行结果存入一个字典实例中。

3.2 自动化批处理

本文所设计的面向网络自动化的高并发插件系统以多线程方式实现特定作业函数在一组主机上批量执行过程见图 3，其内部信息流程如图中箭头所示。下面从变量含义、嵌套作业执行流程、事件监听及非嵌套作业执行流程等不同层面自动化的批处理过程进行阐述。

第一，关于变量含义作以下阐述。作业函数是用于完成特定设备操作管理的、可在远程设备上被执行的、且由用户自定义的一段代码。图 3 以 python 方式定义非嵌套 (no_nested_job) 和嵌套 (nested_job) 两种作业函数，其中非嵌套函数用于执行简单的配置管理任务，而嵌套函数用于执行复杂的配置管理任务。后者将复杂任务拆分为两个简单任务 my_first_job 和 my_second_job。作业管理对象 task 是对作业函数的封装，task 对象向用户提供 start()、

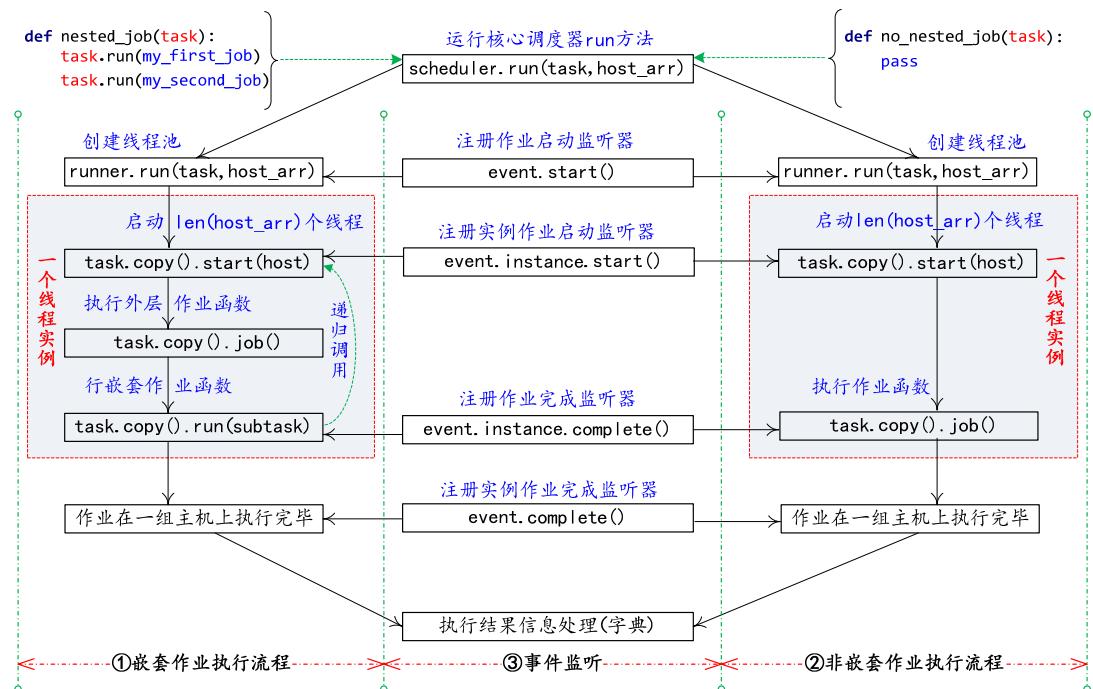


图 3 自动化批处理过程

run()及 copy()等方法，而作业函数本身作为 task 的属性对外暴露。调度器对象 scheduler 的 run()方法是该插件系统为用户提供的编程接口，其参数为封装作业函数的管理对象 task 和设备处理插件程序的结果输出 host_arr。高并发处理对象 runner 是基于 python 实现的线程管理插件类的实例，其提供的 run()方法为用户作业函数的高并发处理提供基础支撑。event 是以事件管理插件类对象为元素的数组引用，为作业函数的执行结果采用事件驱动方式进行处理提供了坚实基础。

第二，关于嵌套作业执行流程作以下阐述。当用户向该插件系统输入一个特定的嵌套作业 nested_job 和一组网络设备 host_arr 后，其内部自动化批量处理流程如下：(1) 将用户自定义的作业函数 nested_job 封装为 task 对象并和设备数组 host_arr 一并作为参数提交给 scheduler 的 run()方法，从而完成自动化批处理任务的启动；(2) 在 scheduler.run()内部，依据配置动态加载相应线程管理插件类并将其实例化 runner 对象，调用 runner.run()方法使作业函数执行进入高并发模式，实际线程开启数量视 host_arr 数据大小而定；(3) 为降低线程间的耦合度，在每一个线程内部，通过 task 对象的拷贝调用其 start()方法，从而在 start()方法内部处理逻辑中完成对外层作业函数的调用；(4) 通过 task 对象的拷贝实现对 task.run()方法的调用，在 task.run()内部采用自动生成的 subtask 对象实现对嵌套在内层的子作业 my_first_job 或 my_second_job 的封装；(5) 为最终完成对嵌套在内层的子作业在特定主机上的执行，在 task.run()方法内部递归调用 subtask.start()方法，直至一个线程内的所有子作业执行完毕；(6) 嵌套作业 nested_job 在 len(host_arr) 个网络设备上以并发模式被执行，从而实现自动化批处理的设备配置管理理念。与嵌套作业执行流程相比，非嵌套作业执行流程中少了递归调用的环节。

第三，关于事件监听作如下两个层次的阐述。(1) 从作业函数执行的流程方面而言，事件监听分为面向一组网络设备的监听和面向一个网络设备的监听。譬如，event.start()在所有网络设备尚未执行特定作业函数前注册；event.instance.complete()在某一个网络设备上执行完特定作业函数后被注

册。(2) 关于事件监听器的驱动方式表述为：在插件系统内部的编程实现中，采用以接口方法调用的形式(譬如 event.complete())依次在 scheduler.run()方法中，在 task.start()方法中以及在 task.run()方法中设置锚点；在插件系统外部，用户可根据自身需求对事件处理插件接口进行编程实现，从而实现一旦锚点中的方法被调用，则以事件驱动的方式引发用户期望的固定动作被自动执行，譬如将作业函数的输出结果进行某种特殊处理后存档。

3.3 高并发模型

高并发特性是衡量一个网络自动化运维工具在实际生产网络中能否应用成功的关键指标。本文设计的高并发线程模型采用 python 线程池的方式实现，详见图 4^[10]。图 2 中的并发处理对象 runner 可看作类 ThreadedRunner 的实例化对象，其属性 num_workers 表征线程池的大小，run()方法的两个参数 task 及 host_arr 的具体含义与前文表述含义一致，待特定任务在一组网络设备上执行完毕后其运行结果存入字典数据结构 AggregatedResult。

```
class ThreadedRunner:  
    def __init__(self, num_workers: int = 20) -> None:  
        self.num_workers = num_workers  
    def run(self, task: Task, host_arr: List[Host]) -> AggregatedResult:  
        result = AggregatedResult(task.name)  
        futures = []  
        with ThreadPoolExecutor(self.num_workers) as pool:  
            for host in host_arr:  
                future = pool.submit(task.copy().start, host)  
                futures.append(future)  
        for future in futures:  
            worker_result = future.result()  
            result[worker_result.host.name] = worker_result  
        return result
```

图 4 高并发线程模型

图 5 所示的高并发作业执行模型是对图 4 所示高并发线程模型的可视化呈现。下面以图 5 中定义的作业函数为例来描述多线程环境下的高并发特性。(1) 当 runner.num_workers=1 时，作业函数 grouped_jobs 运行在单线程模式下，执行时间 T1=3.1+3.2+…+3.(N+N)+3.(NUM_HOSTS)，其中 N 等于线程池大小 num_workers，NUM_HOSTS 等于设备数组大小 len(host_arr)。(2) 当 runner.num_workers>1 时，作业函数 grouped_jobs 面向每个单独网络设备运行在多线程模式下，其内部的两个子作业函数 my_first_job 和 my_second_job 则运行在

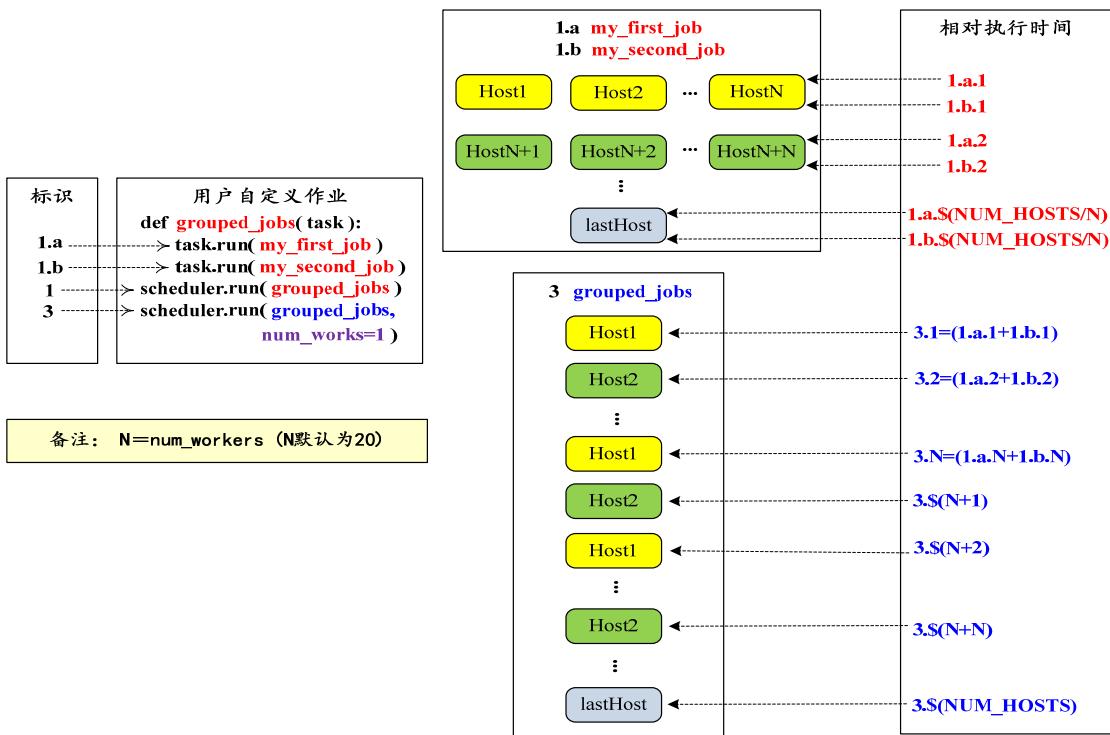


图 5 高并发作业执行模型

单线程模式下，执行时间 $T_2 = (1.a.1 + 1.b.1) + (1.a.2 + 1.b.2) + \dots + (1.a.\$(\text{NUM_HOSTS}/N) + 1.b.\$ (\text{NUM_HOSTS}/N))$ 。 (3) 假设 `my_first_job` 在所有网络设备上执行时间都相等且记作 t_1 , 同理 `my_second_job` 相应时间记作 t_2 ; $N=\text{num_workers}=20$; $\text{NUM_HOSTS}=\text{host_arr}=500$ 。则将上述参数代入公式 T_1 和 T_2 ，则 $T_2/T_1*100\%=[(t_1+t_2)*25/(t_1+t_2)*500]*100\% = 5\%$ ，由此可见在上述条件下，高并发插件系统相较于普通单线程状态下的执行效率提高了 95%，而且这种效率的提高会随着网络规模增加和服务器性能提升变得更加显著。

4 插件系统应用

通常将基于 SSH 协议登陆到远程设备所进行的一系列网络配置管理活动称之为初级网络自动化，而将采用 netconf 或 RestConf 等方式进行的一系列网络配置管理活动称之为高级网络自动化。本文所设计实现的的网络自动化系统基于插件方式架构，因此用户可根据其提供的连接管理插件接口标准编程实现不同的连接管理插件类。常用的连接管理插件类既包括基于 paramiko、netmiko、scrapli 及 pexpect 等 python 库实现的初级网络自动化工具

类，也可以基于 netconf 协议实现高级网络自动化工具类。

为清晰阐述该高并发插件系统的应用场景及其运行性能，文章利用公司部署的华为私有云平台和开源 GNS3 模拟软件构建出包含 500 台网络设备的计算机网络。具体思路简述为：(1) 在私有云中创建 10 台预装 ubuntu 操作系统的虚拟机；(2) 在每台 unbutu 虚拟机上部署一套 GNS3 服务端软件并通过 GNS3 的 GUI 界面或 Web 方式实施控制；(3) 在每台 GNS3 服务器上创建如图 6 所示的包括 500 台模拟交换机的网络拓扑并通过网云节点与真实网络互通；(4) 正确配置模拟网络使主控计算机(同时安装插件系统及 Ansible)可访问 GNS3 网络拓扑中的所有模拟交换机。(5) 在硬件性能方面，虚拟机 CPU 为 16Core，内存为 16 GB，中控计算机的 CPU 为 64 核，内存为 128 GB。

通过实验应用表明：本文设计实现的高并发插件系统作为一个只针对网络自动化运维的轻量级工具，在网络规模较大时其运行效率要明显高于侧重于系统自动化运维的重量级工具 Ansible，并能同时满足初级自动化和高级自动化的实际应用需求。

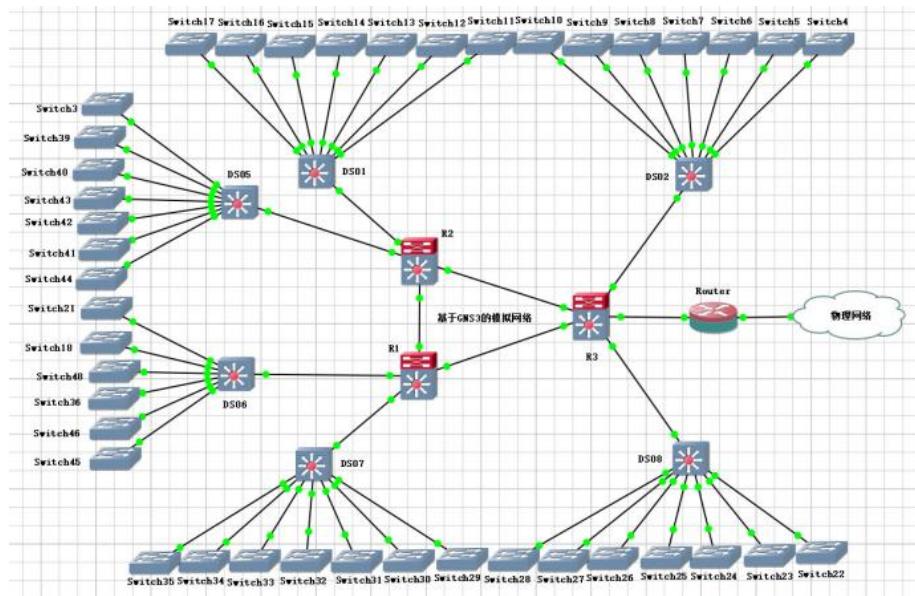


图 6 基于 GNS3 的实验环境

5 结束语

在网络管理日益复杂和 SDN 技术引起网络体系结构变革的大背景下,网络自动化技术成为构建智简网络的必由之路。然而,传统自动化运维工具在应用前期更侧重于面向服务器的操作系统运维及构建于其上的应用软件的配置操作,往往在面向网络自动化运维方面因其系统架构复杂和网络插件匮乏等因素而表现欠佳。因此,本文在吸取 Ansible 及 Salt 等经典工具在自动化软件架构方面成功经验的基础上,提出一种基于插件开发及运行模式的、高并发、高可靠的网络自动化平台架构方法,并在 python 生态下将其编程实现为一种面向网络自动化的高并发插件系统。通过对该软件在依托 GNS3 构建的模拟网络中相关性能数据的对比分析和应用实践,表明基于该方法的自动化工具相较于传统工具而言,其在灵活性、可扩展性和执行效率等方面优势明显,具有良好的应用前景和推广价值。

参考文献:

- [1] EDELMAN J, LOWE S S, OSWALT M. Network programmability and automation[M]. 2018: 10-11. Nanjing: Southeast University Press, 2018: 10-11.
- [2] BARROSO D. Welcome to nornir's documentation [EB/OL]. (2022-01-30). <https://nornir.readthedocs.io/en/latest/index.html>.
- [3] CLAISE B, CLARKE J, LINDBLAD J. Network programmability with YANG[M]. New York, USA: Pearson Education, Inc., 2019: 23-25.
- [4] EDELMAN J, LOWE S S, OSWALT M. Network programmability and automation[M]. Nanjing: Southeast University Press, 2018: 359-360.
- [5] Ansible project contributors. Ansible documentation [EB/OL]. (2022-02-07). <https://docs.ansible.com/ansible/latest/index.html>.
- [6] EDELMAN J, LOWE S S, OSWALT M. Network programmability and automation[M]. Nanjing: Southeast University Press, 2018: 357-359.
- [7] StackStorm project contributors. StackStorm documentation[EB/OL]. (2022-02-07). <https://docs.stackstorm.com/>.
- [8] HOCHSTEIN L. Ansible: up and running [M]. Graveston Highway North, Sebastopol, United States of America: O'Reilly Media, Inc., 2015: 32-35.
- [9] LEVKIVSKYI I, LEHTOSALO J, LANGA L. Protocols: structural subtyping (static duck typing)[EB/OL]. (2022-01-30). <https://www.python.org/dev/peps/pep-0544/>.
- [10] BARROSO D. Nornir API documentation(nornir.plugins.runners)[EB/OL]. (2022-01-30). https://nornir.readthedocs.io/en/latest/api/nornir/plugins/runners/__init__.html.

城域光传送网物理层安全威胁与防护策略分析

何清泉，郭纪炳，熊宗炬，岳清华，骆施宇

中国西昌卫星发射中心，四川 西昌 615000

摘要：通过详细研究光传送网（OTN）物理层当前面临的安全威胁以及相关防护技术，从多个维度对各类防护技术进行横向对比评价，最终探讨并制定了多层次的防护策略。为深入理解系统面临的安全威胁，把窃听、干扰、篡改三类常见的攻击类别，从攻击原理和实现方法等角度进行概述。为阐明光传送网现有的安全防护技术，将环境综合防护、光信号加密、光信号监测、认证技术等防护手段的技术特点和应用现状进行总结。以分析对比和排名的方式，从技术成熟度、防护效果、实施成本等方面，完成对各类防护技术的分析统计和排名，同时结合现网中的实际情况和可能面对的安全防护需求，提出了一种基于多种防护技术的城域光传送网物理层安全策略。

关键词：OTN；安全防护；光纤加密；光信号认证

文献标志码：A

中图分类号：TP311.11

Security Threats and Protection Strategies of Metropolitan Area Optical Transport Network Physical Layer

HE Qingquan, GUO Jibing, XIONG Zongju, YUE Qinghua, LUO Shiyu

Xichang Satellite Launch Center, Xichang, Sichuan 615000, China

Abstract: Through a detailed study of the current security threats and related protection technologies faced by the optical transport network (OTN) physical layer, the horizontal comparison and evaluation of various protection technologies are carried out from multiple dimensions, and finally a multi-level protection strategy is discussed and formulated. In order to deeply understand the security threats faced by the system, three common types of attacks, eavesdropping, interference and tampering, are summarized from the perspective of attack principle and implementation method. In order to clarify the existing security protection technology of optical transmission network, the technical characteristics and application status of environmental comprehensive protection, optical signal encryption, optical signal monitoring, authentication technology and other protection methods are summarized. By means of analysis, comparison and ranking, the analysis, statistics and ranking of various protection technologies are completed from the aspects of technology maturity, protection effect and implementation cost. At the same time, combined with the actual situation in the current network and the possible security protection needs, a physical layer security strategy of metropolitan optical transport network based on various protection technologies is proposed.

Key words: OTN; safety protection; optical fiber encryption; optical signal authentication

城域光传送网基于光纤通信技术，利用光信号在光纤中高速传播的特性，实现数据的高效传输。光传送网的物理层通常由光收发器、光纤、光放大器、光开关等多种精密设备组合而成，其物理层主要承担光信号的精准发送、灵敏接收和稳定传输等

重要工作。在光传送网中，上层的数字信息转化为光信号后，通过光纤这一高速通道传至目的地，再进行光电转换，后续在传输设备中进行光信号放大、整形、滤波等精细处理，以确保光信号的质量始终处于优良状态，为整个网络的稳定运行奠定坚

作者简介：何清泉（1997—），男，工程师，研究方向为光纤通信、信息安全。

实基础。

光传送网凭借自身卓越的大带宽承载能力、低时延传输特性等优势，已成为现代通信网络的核心支柱。然而，光传送网在信息安全方面的防护技术和成果运用还比较滞后，其网络安全性与使用规模相比存在明显差距。在网络攻击手段日益多样化、智能化的背景下，光传送网在传输层面的安全隐患将会逐渐暴露。城域光传送网作为通信承载网络的根基，其安全性直接关系到整个网络的稳定与可靠，影响 OSI 模型上层各类业务的通信质量和通信安全，一旦出现安全问题，将会给各个行业领域带来重大损失。本文致力于深入分析城域光传送网面临的安全威胁，制定行之有效的防范策略。

1 城域光传送网物理层面临的安全威胁

城域光传送网络在物理层上面临的安全威胁主要有三种类型：窃听攻击、干扰攻击、篡改攻击。

1.1 窃听攻击

1.1.1 光纤窃听

由于城域光传送网需要在长距离上进行光信号传输，大多会出现光缆线路暴露在防护等级不高的环境中的情况，在复杂的外部环境中，攻击者可以通过隐蔽手段，在光纤上安装专业窃听设备，在网络使用者正常通信的同时，悄然窃取光信号中所承载的信息。光纤窃听的方法主要有弯曲光纤窃听和光纤分路窃听。

弯曲光纤窃听是将光纤弯曲到一定程度，改变光信号的原始传输方向，迫使部分光信号泄露，再使用具有高灵敏度的光接收器接收泄露出的光信号，从中获取光传送网中传输的信息。

假如弯曲前这段光纤所携带的光功率为 P_1 ，弯曲后这段光纤所携带的光功率为 P_2 ，光纤弯曲功率泄漏系数定义为： $\alpha = (P_1 - P_2) / P_1$ ， α 越大，功率泄露越严重。假定弯曲功率泄漏模型是一圆形截面的光纤，并作弱导光纤的假设，带入光纤弯曲功率泄漏系数公式：

$$\alpha = \frac{2ak^2 \exp[2\gamma a - 2/3(\gamma^3 / \beta^2)R]}{e \sqrt{\pi\gamma RV^2}}$$

式中， a 为纤芯半径， k 为导模径向相位常数， γ 为导模径向衰减常数， β 为直波导的传播常数， R 为光纤弯曲半径， V 为归一化频率，对于一般单模

光纤 $e_v = 2$ 。由计算可知，光纤弯曲功率泄漏系数是由光纤的物理参数和光纤光信号波长等因素共同决定的，而功率泄漏系数的大小决定泄漏功率的大小，泄漏功率的大小决定了光信号窃取能否实现。所以，当物理参数条件合适时，光纤弯曲的泄漏功率足够光接收机正常工作，就可以通过这种手段窃取到光传送网中传输的信息。

光纤分路窃听则是在光纤中插入分路器，将光信号进行分路，把分出来的光信号接入光接收机进行接收，以达到窃取信息的目的。

1.1.2 设备窃听

光收发器、光电转换器等设备作为光传送网中的关键节点，也可作为窃听攻击的选择目标。通过多种手段对光传送网中的设备发动攻击，进而窃取其中传输的数字信息。窃听方法可以分为软件攻击和硬件攻击两种。

软件攻击是通过攻击传输设备的网管、数据库等控制软件，从系统管理层面获取其控制权限，通过非法授权窃取光传送网中承载的业务信息。硬件攻击则是针对光收发器、光电转换器等器件的芯片、电路板等组件进行控制和窃听，进而达到窃取器件中光信号信息的目的。

1.2 干扰攻击

1.2.1 频率干扰

通过发射与光传送网中光信号频率相同或相近的光信号，可以干扰光信号的传输。使用频率干扰引起收发两端之间传递的光信号频率发生偏移，导致接收端无法正确接收光信号，从而影响正常通信。

设原始光信号频率为 f_0 ，干扰光信号频率为 f_1 ，受干扰后光信号频率变为 f ，根据频率叠加原理， $f = f_0 + \Delta f$ ，其中 Δf 为干扰引起的频率偏移量，可表示为 $\Delta f = \frac{P_1}{P} \times (f_1 - f_0)$ ， P_1 为干扰光信号功率， P 为正常光信号功率。

根据波的叠加原理，当干扰光与信号光频率相同、振动方向相同、传播方向相同，则会产生干涉、衍射现象；频率相同、振动方向垂直、传播方向相同，产生偏振光现象；振幅相同、频率相同、振动方向相同、传播方向相反，则产生驻波现象。通过加入相同或相近干扰光源，导致振动矢量的变化，

对原光信号造成影响，达到信号干扰的目的。

1.2.2 相位干扰

通过发射和光传送网中光信号相位相同或相反的光信号，实施相位干扰。相位干扰会使光信号的相位发生偏移，同样会导致光信号无法被正确接收，破坏光信号传输的稳定性。

设原始光信号相位为 φ_0 ，干扰光信号相位为 φ_1 ，受干扰后光信号相位变为 φ ，则 $\varphi = \varphi_0 + \Delta\varphi$ ，其中 $\Delta\varphi$ 为干扰引起的相位偏移量，可表示为 $\Delta\varphi = \frac{P_1}{P} \times (\varphi_1 - \varphi_0)$ ，其中 P_1 为干扰光的功率， P 为正常光信号功率。

当两个波形相同但相位完全相反的波叠加时振幅为 0；当两个波形相同、相位相同的波叠加的时振幅为原来的 2 倍。所以加入相位相同或相反的干扰光信号后，会导致信号幅度的增强或相消。此干扰方式通过加入干扰光源引起振动幅度的变化，对原光信号造成影响，达到信号干扰的目的。

1.3 篡改攻击

1.3.1 光信号篡改

攻击者可以对光传送网中的光信号进行篡改，以达到攻击、欺骗等目的。光信号篡改的方法主要有光信号插入、光信号删除、光信号替换等。

光信号插入是在光信号中恶意插入虚假、错误的信息，导致接收端接收错误或发生误判。光信号删除是直接删除传输中的部分或全部光信号，破坏信号的完整性。光信号替换则是将光信号中的部分信息替换为虚假的信息，用来欺骗接收端。

1.3.2 设备篡改

攻击者还可以对光传送网中的设备进行篡改，达到破坏设备或控制设备的目的。设备篡改的方法主要有软件攻击和硬件攻击。

软件攻击是通过攻击设备的软件系统，获取设备的控制权限，然后篡改设备的配置信息，使设备按照攻击者的意愿运行。硬件攻击则是针对设备的芯片、电路板等硬件组件进行攻击，篡改设备硬件，导致其正常异常，从而影响设备的正常功能。

2 防护技术

2.1 环境综合防护

多层次、全方位防护地做好设备和光信号所处

环境的安全防护，是保证城域光传送网络通信安全的基础。目前，在机房、设备、光缆的安全防护和配套设施建设等方面，已有详细的标准、规范和成熟的技术、策略。

相关标准有：《GB/T 22239—2019 信息技术网络安全等级保护基本要求》《涉密机房建设标准》《GB 51103—2015 电磁屏蔽室工程施工及质量验收规范》《GB/T 1899—2023 全介质自承式光缆》《GB/T 43556.1—2023 光纤光缆线路维护技术》等。其明确规定了机房和通信系统的安全保护能力要求、涉密机房环境要求、电磁屏蔽设施的建设要求、光缆的质量检验和运输储存要求等内容，为开展相关设备设施建设提供了依据和指导。

相关技术有：防电磁泄漏技术、门禁识别、身份识别、访问控制、入侵检测等。应用这些技术，可以防范电磁信号泄露、限制未授权人员、管控用户行为、及时发现隐藏风险，能够有效提升系统的安全防护能力。

2.2 光信号加密技术

光信号加密技术是为了有效防范窃听而提出的，通过对光信号进行高强度加密，可以使得攻击者即使窃取光信号，也无法解密其中的信息。光信号加密技术属于光学领域的研究成果，通过控制光信号的振幅、相位、偏振态等特征值，对光纤中传输的光信号进行加密，借助光的多自由度设计的密钥，能对光信号进行多维编码，可以发挥光并行处理的优势，实现对光纤中传输的信息进行加密的目的。光层加密直接在光信号传输的过程中进行加密，具有比较高的实时性和可靠性。

光信号加密技术主要有以下三类。

基于光信号干涉的加密技术：在光信号中引入特殊干涉模式，可以实现传输数据的加密。例如，通过干涉光束的相位差来加密信息，解密时只需使用与加密相同的参考光束进行干涉即可恢复原始信息。干涉技术利用了光的相干性和并行处理能力，能够高效地实现数据加密和解密。

基于相位编码的加密技术：通过改变光信号的相位信息来实现数据加密。例如，双随机相位编码技术通过在输入信号和参考信号上分别施加随机相位掩膜，在接收端通过相位的逆变换来恢复原始信息。相位编码技术具有较好的稳定性和可靠性，

适用于需要高安全性的数据传输。

基于光纤传输参数的加密技术：利用光信号的波长、振幅、相位、偏振态等传输特性，进行数据保护。例如，光纤传感器可以通过测量光纤中光的相位变化来实现对物理量的加密测量。这种技术利用了光纤的物理特性和传输特性，能够实现高精度和高安全性的数据加密。

虽然光信号加密具有许多优势，但它的劣势同样比较明显。一是光信号加密技术的成熟度还有待提高，技术对实验条件要求较高，实现的复杂度较大。二是成本较高，大多需要精密的光学器件和昂贵的设备。三是技术的应用程度还不够广，以上两点也在一定程度上导致技术推广不便。

2.3 光信号监测技术

通过对光信号各个特征值的监测，可以及时发现城域光传送网中光信号的异常变化，及时开展针对性防御和处置。监测内容主要是光信号的功率、频率、相位等特征。

光功率监测：采用先进的光功率监测技术，实时监测光传送网中的光功率变化。一旦发现光功率出现异常变化，立即发出警报，以便及时采取相应的措施。可以通过设置合理的光功率阈值，当光功率超出阈值范围时，自动启动应急响应机制。

光信号频率监测：利用频率监测技术，对光传送网中的光信号频率进行实时监测。当光信号频率发生异常变化时，迅速发出警报，提醒相关人员进行处理。可以采用频率锁定技术，确保光信号的频率始终保持在稳定的范围内。

光信号相位监测：借助相位监测技术，实时监测光传送网中的光信号相位变化。一旦发现相位异常，立即采取措施进行调整。可以采用相位补偿技术，对光信号的相位进行实时补偿，确保光信号的相位始终保持正确。

2.4 认证技术

认证技术在传统网络中已经得到了很好的应用和发展，光传送网与认证技术的结合大致可分为直接使用和技术迭代两种，直接使用就是将传统认证技术的设备认证运用到光传送网中，保证设备、用户真实且合法；技术迭代则是结合数字签名、哈希认证等手段，发展出的一种新兴的光信号认证技术。

2.4.1 光信号认证

目前比较成熟的光信号认证技术主要是数字签名认证。通过对光信号进行数字签名，验证光信号的真实性和完整性。在发送端使用私钥，通过签名函数 $Sign(private_key, message)$ 生成一个唯一的签名值附加到原始数据上，对光信号进行签名，其中， $private_key$ 为发送方的私钥， $message$ 为传送的光信号。接收方使用公钥对签名的有效性进行验证，其公式为 $Verify(public_key, message, signature)$ ，验证通过则可以确保数据真实性。

常用的数字签名方式是借助哈希运算，在发送端对光信号进行哈希运算，生成哈希值，传递至接收端后，对接受光信号和哈希值进行验证，确保光信号未被篡改。哈希函数（Hash Function）使用散列算法，可以把任意长度的输入换算成固定长度的输出，即哈希值。常用的哈希函数有：SHA-1、SHA-256、MD5 等。由于哈希运算具有很好的稳定性、高效性、雪崩效应和不可逆性，通过哈希运算对光信号进行数字签名能可靠地检查光信号的真实性和完整性。

2.4.2 设备认证

通信系统中的设备认证技术主要是验证设备身份，根据身份赋予相应的权限，以达到控制访问、操作的目的。运用设备认证技术可以防止未授权访问、防止篡改攻击等威胁。

常用的设备认证技术有数字证书认证、密码认证、硬件认证等。数字证书认证是对设备颁发数字证书作为电子凭证，验证设备的合法性和进行加密通信。密码认证是最常见的认证技术之一，通过密钥验证开展身份认证。硬件认证是基于令牌等物理设备，每次验证时根据时间或使用次数等变量生成一次性密码，用来验证身份并授权访问。

根据认证技术的叠加使用情况，可以将设备认证分为单因素认证、双因素认证和多因素认证，采用双因素认证和多因素认证的安全性更高。

3 安全策略

对光传送网主要的安全防护技术进行综合分析，重点对比各个技术在成熟度、防护效果、成本方面的优劣，对其进行排名并说明理由，详见表 1。

表1 防护技术优劣对比

技术名称	技术成熟度		防护效果		成本优势	
	排名	原因	排名	原因	排名	原因
环境综合防护技术	1	有详尽的标准、规范等文件为依据,应用广泛	3	主要用于确保机房、光缆等设备设施的整体安全,防护效果在广泛的应用中经过反复验证有效	2	其成本相对固定,且在建设时有国家标准进行要求和约束,可划入工程建设的成本,不需要单独考虑
光信号加密技术	5	属于光学加密的范畴,加密、解密所需的条件比较严苛	1	对原始光信号进行加密,能够较好的保护其上承载的数据,能有效防止窃听	5	需要复杂的加密算法和和高精度光学器件作支持,成本较高
光信号监测技术	2	能在收端持续且直观地监测光信号的功率、频率、相位等指标,几乎所有光传送系统设备都支持该功能	5	不直接提供防护功能,协助预警和通报潜在威胁	1	监测相关指标需要相应的传感器和数据系统,但大部分光传送设备内置相关功能,节省了二次建设的成本
光信号认证技术	4	由数字签名、哈希运算等技术作为支撑,是一种比较新的认证技术	4	在收、发两端进行光信号认证,能够识别出光信号的真实性和完整性,防护范围较窄	4	属于新型认证技术,相比传统的认证技术成本稍高
设备认证技术	3	与传统网络中的认证技术基本相同,技术比较成熟	2	对传输设备和用户设备进行认证,可以有效识别设备身份,避免伪装和窃听攻击	3	主要依赖于密码算法和验证机制,需要少部分硬件,总体成本较低

由于各个技术的侧重点不同,其防护效果和成本并非单一因素能够决定,此处的排名并非绝对客观。在实际应用中,城域光传送通常需要结合使用多种防护技术,才能达到多层次、全方位的安全防护效果。

根据防护技术的优劣对比,可以看出,在技术成熟度方面,环境综合防护、光信号监测、设备认证技术都达到了较高的成熟度,应用范围较广;在防护效果方面,光信号加密、设备认证、环境综合防护技术的优势相对较大,能够有效防范各类攻击;在建设成本方面,光信号监测、环境综合防护、设备认证技术的成本相对较低。综合来看,环境综合防护技术在技术成熟度、防护效果、成本三个方面都具有优势,是首要推荐的防护技术;光信号加密技术具有卓越的防护效果,但技术成熟度还有待提升、建设成本高,在实际应用中要认真分析使用的必要性,视实际情况决定是否采用;光信号监测技术成本较低、应用广泛,但不直接提供防护功能,只能作为预警和协助分析的手段;光信号认证技术与其他技术相比在三个方面都没有明显优势,实际应用中,在具有其他防护手段的情况下,可以不考

虑应用该技术;设备认证技术在三个方面都属于中等水平,没有明显的劣势。

根据对城域光传送网防护技术的对比评价,可以制定如下的安全策略:

首先,要依据《GB/T 22239—2019 信息技术网络安全等级保护基本要求》《涉密机房建设标准》《GB/T 1899—2023 全介质自承式光缆》等标准明确实际系统的安全等级,并规范开展环境建设。对设备机房、线缆间等重点场所使用防电磁泄漏技术、门禁识别、访问控制、身份识别、入侵检测等技术进行安全防护。同时还要建立合理可靠的应急响应机制、人员管理制度、安全保密规章等进行管理管控。通过开展环境综合防护建设,为城域光传送网构建一个牢固可靠的外部环境。

其次,根据经费预算,选择合适的防护技术开展应用。经费预算紧张时,可以选择使用光信号监测技术,对光纤中光信号的功率、频率等特征值进行监测,达到实时反馈光信号的异常变化、及时预警的目的,方便发生攻击行为时,及时开展针对性防御和处置。经费充足时,可以选择设备认证技术,对系统中的传输设备和用户设备进行认证,严控各

类设备和用户的权限，能够有效防范攻击行为。

另外，还可以结合传送信息的重要程度，进行全面分析，对于保密要求高的系统，可以采用光信号加密技术进行防范。

运用这一策略，可以有效防范大部分针对城域光传送网的攻击行为，减少系统的安全风险，有效保障系统中传输信息的安全性和可靠性。

4 结束语

当前，信息攻击和窃取技术日新月异，手段层出不穷，城域光传送网在物理层所面临的安全威胁是复杂多样且极具挑战性的。这些安全威胁不仅涉及技术层面，还与管理、策略等多个方面密切相关。只有更深入地理解了这些安全威胁的本质，才能有效应对。想要有效保障城域光传送网的信息安全，必须采取综合的防范措施，从技术应用、管理规范、策略规划等多个角度入手，制定和完善相关保护措施。同时，还应持续研究光传送网的安全问题，以前瞻性的眼光去探讨城域光传送网可能面临的各种威胁，积极探索和应用新的安全技术和防范策略，提高光传送网的整体安全性，这样才能确保光传送网的稳定运行，为信息传输提供坚实的保障。

参考文献：

- [1] 周立. 光纤保密通信系统的设计、实现及性能分析[D]. 南京: 南京邮电大学, 2021.
- [2] 李珊珊. 正交频分复用无源光网络物理层安全防护技术研究[D]. 武汉: 华中科技大学.
- [3] DAHAN D, MAHLAB. SECURITY U threats and protection procedures for optical networks[J]. IET Optoelectronics, 2017(5).
- [4] 陶立康. 无源光网络物理层安全防范策略的研究[D]. 南京: 南京邮电大学, 2013.
- [5] LI Y L, HUA N, LI J D, et al. Optical spectrum feature analysis and recognition for optical network security with machine learning[J]. Optics Express, 2019(17).
- [6] 张瑛瑛. 电力数据调度的光传输网设计与安全应用探究[J]. 通讯世界, 2024(7): 130-132.
- [7] VIPUL A, RAMISETTI A, SIVA G H, et al. An ultrafast all optical encryption decryption scheme based on XOR logic for secure transmission in optical networks[J]. Journal of Physics: Conference Series, 2021(1).
- [8] EDUARDO M, SOARES D A. 100 Gbit/s AES-GCM cryptography engine for optical transport network systems: architecture, design and 40 nm silicon prototyping[J]. Microelectronics Journal, 2021.
- [9] 魏睿, 包燕军. 基于超构表面的多维光信息加密[J]. 中国激光, 2023(18): 54-68.
- [10] 吴流丽, 廖建华, 汪文晓, 等. 光传输网安全风险分析及防御技术研究[J]. 信息安全与通信保密, 2023(10): 60-66.
- [11] BAI L R. Optical service unit (OSU)-based next generation optical transport network (NG OTN) technology and verification[C]//MATEC Web of Conferences, 2021.
- [12] 王祥青. 光网络物理层安全认证及加密技术研究[D]. 北京: 北京邮电大学, 2021.
- [13] YIN A H, LI Q, ZHU M. Secure authentication scheme for 10<ce:hsp sp="0. 25"/>Gbit/s Ethernet passive optical networks[J]. Optik - International Journal for Light and Electron Optics, 2014(20).
- [14] MARIJA F, CARLOS N, ANDREA D G, et al. Optical network security management: requirements, architecture, and efficient machine learning models for detection of evolving threats [Invited][J]. Journal of Optical Communications and Networking, 2021(2).
- [15] 吴坤. 基于光纤信道特性的无源光网络物理层安全技术研究[D]. 北京: 北京邮电大学, 2023.
- [16] FAN C, YANG H Z, HE J L, et al. OTN technology development and application in metropolitan electronic Power Network[J]. Journal of Physics: Conference Series, 2021(4).

基于 Nornir 的网络自动化运维平台构建

李朝阳，熊宗炬，张小锋，周 波，岳清华

中国西昌卫星发射中心，四川 西昌 615000

摘要：通过对传统自动化运维工具在网管领域的应用瓶颈进行研究，设计出一种基于 Nornir 构建多用途、高并发、轻量级、可编程的网络设备自动化运维平台的技术解决方案。该方案在深入分析 Nornir 源码的基础上，针对平台架构设计、资产管理插件开发和连接管理插件开发等关键技术环节的实现给出一般性方法。该自动化平台在实验环境中的一系列应用表明，提出的方案在应对日益复杂的网络管理运维工作中优势明显，具有良好的推广应用价值。

关键词：Nornir；网络自动化；插件开发；SSH；Paramiko

文献标志码：A **中图分类号：**TP311.11

Construction of Network Automation Operation and Maintenance Platform Based on Nornir

LI Zhaoyang, XIONG Zongju, ZHANG Xiaofeng, ZHOU Bo, YUE Qinghua

Xichang Satellite Launch Center, Xichang, Sichuan 615000, China

Abstract: After studying the application bottleneck of traditional automatic operation and maintenance tools in the field of network management, this paper designs a solution to build a multi-purpose, high-concurrency, lightweight, programmable network automatic operation and maintenance platform based on Nornir. On the basis of in-depth analysis of Nornir source code, this solution provides general methods for the realization of key technology links such as platform architecture design, inventory management plug-in development, and connection management plug-in development. A series of applications of the automation platform in the experimental environment show that the solution proposed in this paper has obvious advantages in dealing with the increasingly complex network management operation and maintenance work, and has good promotion and application value.

Key words: Nornir; network automation; plugin development; SSH; Paramiko

出于简化网络架构、实现业务快速灵活部署等方面考量，在 SDN（software definition network）时代，实现网络大规模配置管理、网络性能数据快速收集、网络成规模迁移、网络运行报告生成以及网络故障快速定位等运维操作的自动化已成为未来网络运维领域的主流技术趋势^[1]。换言之，通过工具实现网络自动化部署、运行和运维，逐步减少对“人”的依赖。一些曾经只关注系统运维层面的传统自动化工具亦逐步将其功能向网络自动化领域扩展^[2]，典型代表有 Ansible、Salt 及 Stackstorm 等。传统自动化运维工具在网络自动化领域的应用通常会遭遇插件工具不完善、程序调试不方便、学

习曲线较陡峭及国产设备不适应等方面的问题。因此，本文提出一种基于 Nornir 构建多用途、高并发、轻量级、可编程的网络设备自动化运维系统的技术解决方案，以适应日益复杂的网络管理运维工作新情况。

1 关键技术

1.1 Nornir

Nornir 是一个纯 python 语言开发的面向网络设备管理的插件式自动化框架^[3]，旨在解决传统自动化框架调试诊断工具匮乏及系统整合困难等问题。Nornir 的体系结构如图 1 所示，包括 Nornir

基础类库、插件接口、插件程序、调度器及用户接口五部分。

Nornir 基础类库为调度器主要功能的实现、各类插件程序的二次开放及任务函数的封装执行提供基础支撑环境,重点包括如图 1 所示的 7 类基础类库。其中, task.py 模块的主要作用是对各类任务函数提供封装工具,实现任务函数执行过程的自动调度和执行结果的自动处理。inventory.py 模块为设备资产资源的解析处理提供方法和工具,其最重要的类是 Host 类。Host 类向用户暴露设备连接方法、设备资产操作方法及设备连接关闭方法等。模块 filter.py 用于配合 inventory.py 模块工作,为设备过滤和选定提供方法及工具。register.py 模块一方面为 Nornir 的各类插件程序二次开发提供标准接口;另一方面为 Nornir 的各类插件程序合法加载提供自动注册和手动注册方法。configuration.py 模块为 Nornir 各类参数解析提供基础方法,尤其是连接参数、各类 yaml 配置文件及日志记录相关参数等。

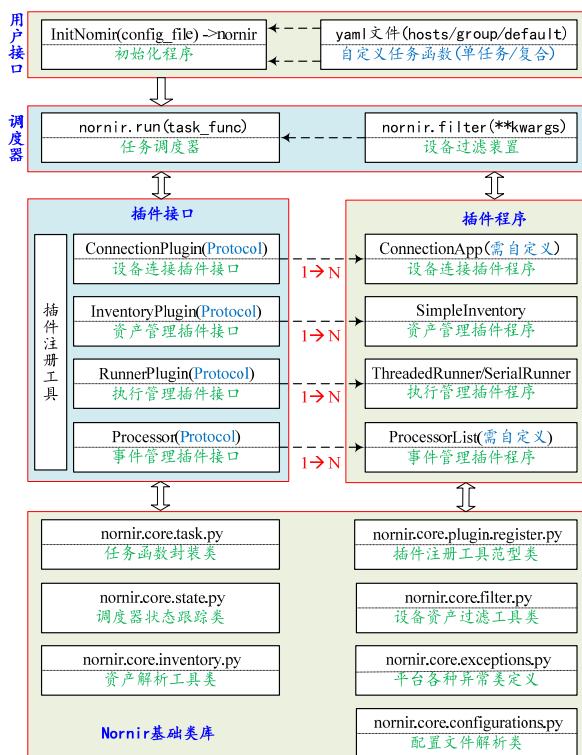


图 1 Nornir 体系结构图

特定插件接口类通过继承 typing_extensions.Protocol 类^[4]从而为相应插件程序的编码实现提供标准规范。Nornir 向用户提供资产管理插件接口、

设备连接插件接口、执行管理插件接口和事件管理插件接口四个标准规范,Nornir 用户可根据自身实际需求定义多个特定功能的插件程序。默认情况下,Nornir 平台提供名为 ThreadedRunner 和名为 SerialRunner 的执行管理插件程序和名为 SimpleInventory 的资产管理插件程序。Nornir 通过不同插件程序的二次开发使整个平台功能具有良好的可扩展性,以连接插件管理程序实现为例,用户可基于 paramiko、scrapli 及 netconf 等不同设备连接方式编码实现相应的设备连接管理插件程序,只要实现 open()方法、close()方法以及实例化 connection 属性即可。

用户接口一方面通过初始化函数 InitNornir()加载各类 yaml 配置文件从而提取设备信息及其连接资源,另一方面将用户自定义的任务函数交由 Nornir 实例进行封装从而为任务执行做好准备。将用户借助设备过滤装置所获得的一组网络设备交由 InitNornir()函数返回的调度器实例对象,从而通过 Nornir 实例对象的 run()方法实现面向大规模设备配置管理的自动化批处理操作。

1.2 设备连接方法

Nornir 的灵活性和可扩展性来源于其统一规范的设备连接管理插件接口。在 python 生态内,Nornir 用户可依据 Pexpect、Paramiko、Netmiko、NAPALM、Scrapli、Telnetlib 及 ncclient 等多种类库工具构建适合自身实际需求的 Nornir 设备连接管理插件程序。各种依赖库的底层依赖协议及相应特点描述见表 1 所示。

1.3 配置模板

在网络编程与自动化的背景下,模板的应用对于降低因人为因素而导致网络配置错误出现的概率、保持大规模网络配置管理的一致性以及节省网络配置管理的时间成本等方面具有非常显著的优势^[14]。通常采用特定的模板语言构建与实际应用环境相符的模板应用,最典型的模板应用出现在 Web 应用程序的开发中。同时,模板语言一定会和具体的编程语言相配合而动态高效地产生各种模板应用。正如 golang 语言和 Ruby 语言有其内建的模板语言一样,python 生态中最常用的模板语言是 django 框架自带的模板语言和 jinja2。在本文所设计的基于 Nornir 的自动化网络运维平台中采用

表 1 网络设备连接方法

工具名称	依赖协议	特点概述
Pexpect	Telnet/SSH	Pexpect 是一个纯 python 模块，用于生成子应用、控制子应用以及对输出中的预期模式做出相应,但对 windows 平台支持有限 ^[5]
Paramiko	SSHv2	Paramiko 是 SSHv2 协议的 Python 实现, 它简化了主机和远程设备之间的 SSHv2 交互, 同时提供客户端及服务端操作 ^[6]
Netmiko	SSHv2	基于 Paramiko 库开发而成, 并在 Paramiko 基础上面向多厂商设备的自动化配置管理进行了一系列的简化操作 ^[7]
NAPALM	SSHv2	NAPALM 是一种为多厂商网络设备提供统一 API 编程接口的 python 库, 它依赖于 netmiko 但多华为及华三设备支持度欠佳 ^[8-9]
Scrapli	Telnet/SSH	Scrapli 是面向网络设备提供 SSH 及 Telnet 协议应用的 Python 库, 其连接速度快于 Paramiko 及 Netmiko 等, 且支持 OpenSSH ^[10]
Telnetlib	Telnet	Telnetlib 是 python 的内建模块, 用于对远程设备实施控制管理, 鉴于 Telnet 协议的安全性, 它在生产网络中很少使用 ^[11]
ncclient	Netconf	ncclient 应用于采用网络编程方式对远程网络设备进行配置和管理, 它是 netconf 协议客户端的一种 python 实现, 属于高级网络自动化范畴 ^[12-13]

jinja2 模板语言作为网络配置模板应用的主要工具, 即考虑到 python 在 jinja2 模板渲染方面的天然优势, 也考虑到诸如 Ansible 及 Salt 等自动化运维工具在 jinja2 方面的成功应用实践^[15]。

2 平台构建

2.1 平台总体架构设计

基于 Nornir 的网络自动化运维平台总体架构设计见图 2, 整个设计围绕 Nornir 框架展开, 主要包括插件程序二次开发、任务函数及其配套模板开发及用户自定义配置文件三部分。

Nornir 的各类插件程序二次开发的丰富程度直接影响着整个自动化运维平台的应用推广前景。

Nornir 向用户开放 InventoryPlugin、ConnectionPlugin、RunnerPlugin 及 Processors 四个插件程序二次开发接口, 分别对应于资产管理、连接管理、执行管理及事件处理。

对于资产管理插件程序的二次开发, 需重写 InventoryPlugin 接口的 load()方法。Nornir 默认情况下向用户提供了 SimpleInventory 插件程序, 其实现过程包括两个步骤: (1) 将配置文件 hosts.yaml、groups.yaml 及 defaults.yaml 中的资产信息转换为 Hosts 类实例、Groups 类实例及 Defaluts 类实例; (2) 以之前产生的 Hosts 类实例、Groups 类实例及 Defaluts 类实例作为参数构建资产管理工具类 Inventory 的实例并返回。在面向大规模及多厂商网络设备的自动化运维中, 如果采用基于 yaml 文件的 SimpleInventory 插件程序管理各类资源信息势必难以满足实际需求。因此, 可在 SimpleInventory 的基础上适当改变设备资源信息的存储方式即可满足大规模及多厂商的应用需求, 并由此而衍生出诸如 MysqlInventory、CSVInventory 及 ExcelInventory 等不同资源管理插件程序。

对于连接管理插件程序的二次开发, 需重写 ConnectionPlugin 接口的 open()方法及 close()方法。新版 Nornir 默认情况下不提供任何设备连接管理插件程序^[16]。正因如此, 用户至少可基于 Pexpect、Paramiko、Netmiko、NAPALM、Scrapli、Telnetlib 及 ncclient 等不同工具构建符合实际需求的连接管理插件程序。鉴于当前的网络自动化现状(初级与高级网络自动化会在很长一段时间内并存), 本文基于 Nornir 构建的网络自动化运维平台即可满足基于 SSH 协议的初级网络自动化, 也可满足以 Netconf 为代表的高级网络自动化, 可同时应用于普通园区网以及数据中心网络的运维管理。

对于执行管理插件程序的二次开发, 需重写 RunnerPlugin 接口的 run()方法。Nornir 默认提供 SerialRunner 及 ThreadedRunner 两个执行管理插件程序, 其中, SerialRunner 采用串行方式在一组设备上执行特定任务函数, 而 ThreadedRunner 则基于 python 线程池采用并行方式在一组设备上执行特定任务函数。一般情况下, 默认配置即可满足实际应用需求。如果用户的硬件条件非常优越, 则可基于 python 进程池采用并行方式在一组设备上执行特定任务函数。

对于事件处理插件程序的二次开发，需重写 Processor 类的 task_started()、task_completed()、task_instance_started()、task_instance_completed()、subtask_instance_started() 及 subtask_instance_completed() 的六个方法。至于每个方法的具体实现逻辑没有具体标准，完全取决于用户的应用需求，重点关注特定任务函数执行前后需要 Nornir 额外完成的事务性工作，譬如输出提示信息和按要求保持配置文件等。

关于任务函数及其配套模板的开发，重点围绕配置和监视两个层面展开。首先，在 Nornir 框架下，对于设备的监视和配置则需通过用户自定义的非嵌套和嵌套任务函数实现，至于嵌套与否则取决于监视配置逻辑的复杂度。其次，对于设备监视任务而言，可采用 ntc-templates 工具提供的丰富模板库将任务函数的返回结果从非结构化形式转换为结构化的 JSON 数据，为网络自动化程度的提高提供必要条件。考虑到 ntc-templates 对华为及华三设备缺乏支持，可采用 TextFSM 工具进行相应解析模板的二次开发。最后，对于设备配置任务而言，

可采用 Jinja2 开发适合不同应用场景的设备配置模板，同时借助配置管理数据库 CMDB 或存储为其他形式的资源文件对配置模板进行准确按需渲染，从而为大规模多厂商的自动化配置场景提供技术保证。

2.2 资产管理插件实现

Nornir 框架中对其将要进行配置与监视的所有网络设备及其附加信息统称为资源。Nornir 通过 Hosts 类、Groups 类及 Defaults 类对被管资源进行抽象刻画，并进一步基于其内置 Inventory 容器类对所有资源进行封装同时向外提供过滤工具，其结构见图 3。了解资产容器类 Inventory 的内部结构对于资产管理插件程序的二次开发至关重要。因此，关于该结构图本文阐述如下：(1) 资产容器类实例对象与执行管理类实例对象一样作为 Nornir 调度器的关键属性而存在，在任务函数执行过程中起到提供设备数组的作用；(2) 资产容器类拥有三个类型分别为 Hosts、Groups 及 Defaults 的重要属性，为设备资源的提取、筛选及分组等操作提供方便；(3) 各个基础类之间的继承关系见图 3 所示，

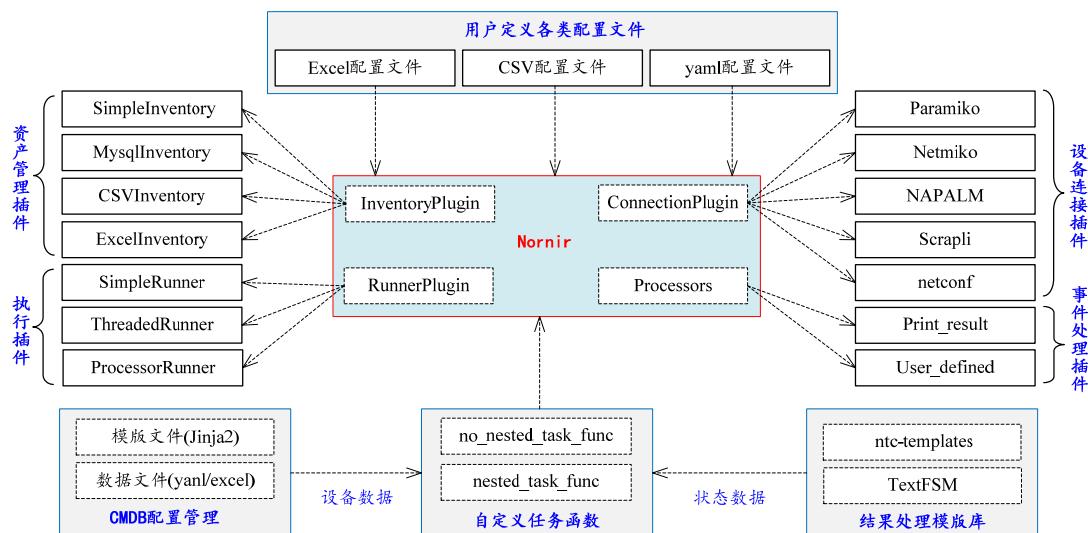


图 2 平台总体架构

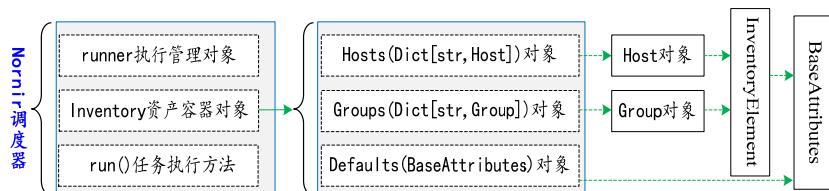


图 3 资产容器类结构

各类常见属性包括用户名、密码、SSH 端口、设备 IP 地址及操作系统类型等。

为保证资产容器类能够准确掌握用户以不同形式提供的设备资源（譬如 yaml 文件或 Excel 文件），必须按照 Nornir 的 InventoryPlugin 接口所定义的标准开发相应的资产管理插件程序（类），即重写其 load()方法。本文在对 Nornir 默认提供的名为 SimpleInventory 的资产插件管理程序源代码进行深入分析的基础上，设计出通常情况下重写资产管理插件类的 load()方法的基本流程框图，详见图 4。

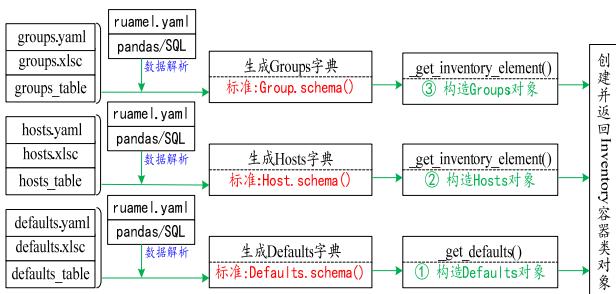


图 4 load 方法实现示意图

本文关于 load()方法重写的一般思路阐述如下：(1) 重写的目的在于将设备资产信息转换为资产容器类可以识别的 Hosts 对象、Groups 对象和 Defaults 对象。(2) 无论采用 yaml 文件、Excel 文件或数据库中的哪一种形式向 Nornir 提供设备资源信息，其最终形式都必须能够转换为 python 字典数据并符合相应的数据模型。以 hosts 资源信息为例，对生成的 Hosts 字典要与 {"sw01":host_obj} 结构相似，其中 sw01 表示设备名称，host_obj 是 Host 类的实例，host_obj 结构应与图 5 所示数据模型相吻合。(3) 与 Hosts 字典类似，Groups 字典及 Defaults 也应符合与之相对于的数据模型。(4) 将各类设备资源文件转换为相应的 python 字典结构数据可根据资源信息存储方式不同而采用 ruamel.yaml 模块、pandas 库或 sql 等多种工具。(5) 为创建并返回资产容器类 Inventory 的实例对象，必须将解析得到的 Defaults 字典、Hosts 字典及 Groups 字典分别转换为 Inventory 类可接受的 Defaults 对象、Hosts 对象及 Groups 对象。(6) 转换过程中可采用默认资产管理插件程序 SimpleInventory 的 _get_defaults()方法和 _get_inventory_element()方法。(7) 由此可见，资产插件管理程序（类）实现的

重点在于通过适当的数据解析工具将资源文件解析生成符合 Inventory 类规范的 python 数据字典，至于编码实现可参考 nornir_table_inventory 资源管理插件程序。

```

from nornir.core.inventory import Host
import json
print(json.dumps(Host.schema(), indent=4))

{
    "name": "str",
    "connection_options": {
        "$connection_type": {
            "extras": {
                "$key": "$value"
            },
            "hostname": "str",
            "port": "int",
            "username": "str",
            "password": "str",
            "platform": "str"
        }
    },
    "groups": [
        "$group_name"
    ],
    "data": {
        "$key": "$value"
    },
    "hostname": "str",
    "port": "int",
    "username": "str",
    "password": "str",
    "platform": "str"
}
  
```

图 5 与 Host 类匹配的数据模型

2.3 连接管理插件实现

掌握基于连接插件类的任务函数执行流程对于连接管理插件程序(类)的二次开发至关重要。因此，本文在对 nornir、nornir_netmiko 及 nornir_netconf 源代码进行深入分析的基础上，设计出一般情况下基于连接管理插件类的任务函数执行流程框架，详见图 6。整个流程结构分为任务函数逻辑、设备连接管理逻辑和连接插件实现逻辑三个层次。

对于一个特定任务函数而言，其实现过程包括建立连接对象、远程执行命令、构造执行结果和关闭网络连接四个阶段。其中，关闭网络连接阶段是通过调用主机对象的 close_connections()方法实现的，该方法内部通过调用连接插件类的 close()方法网络连接资源关闭操作。同理，建立连接对象阶段则通过调用主机对象的 get_connection()方法实现。该方法有 conn_name 和 config 两个参数，config 为远程设备连接过程提供 ssh 配置参数，conn_name 代表特定连接插件类的别名。conn_name 在设备连接、连接插件注册、任务函数定义及设备连接参数定义等各个环节都非常重要，且务必保证其一致性及唯一性。

get_connection()方法的内部逻辑包括图 6 所示的四个关键步骤：(1) 以 conn_name 为参数通过

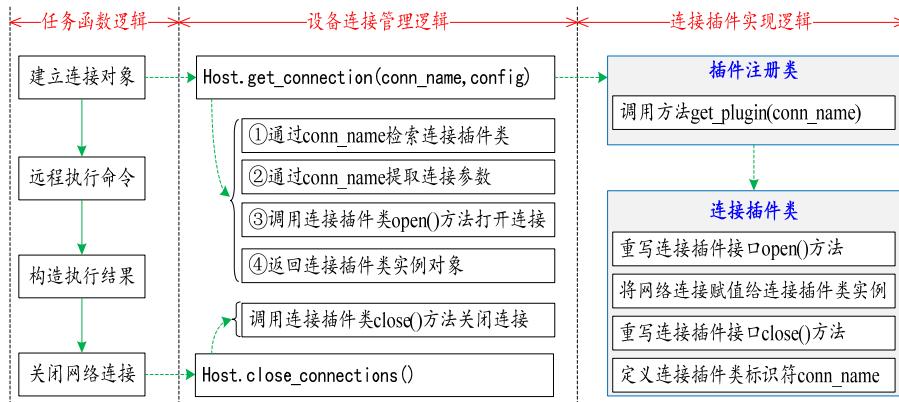


图 6 任务函数执行流程

插件注册类的 `get_plugin()` 方法检索到用户开发实现的连接插件类(该类必须注册);(2)以 `conn_name` 为索引在用户提供的配置文件中解析提取远程连接网络设备所需各类参数;(3)将上一步所获参数传递给连接插件类的 `open()` 方法从而打开执行命令的网络连接对象;(4)返回持有已打开网络连接对象的连接插件类实例。

通过以上对基于连接插件类的任务函数执行流程的分析,可得出连接插件程序(类)二次开发的一般性指导方法:(1)重写连接插件接口 `ConnectionPlugin` 的 `open()` 方法;(2)将上一步中建立的网络连接对象赋值给连接插件类的 `connection` 属性;(3)重写连接插件接口 `ConnectionPlugin` 的 `close()` 方法;(4)给新实现的连接插件程序(类)定义一个全局唯一的标识符 `conn_name`,以适应插件注册、网络连接打开及关闭等操作需求。如果采用 `netmiko` 作为连接管理插件程序二次开发的设备连接工具,则相应连接插件类的 python 实现如图 7 所示。

3 平台应用

3.1 应用场景

本文所构建基于 `Nornir` 的网络自动化网络运维平台主要用于对大规模网络设备进行配置管理和状态监视。应用于网络设备配置管理的一般流程见图 8。首先,根据实际需求基于 `Jiajn2` 模板语言创建相应配置模板。其次,从资源信息存储介质(`Excel` 文件、`yaml` 文件或配置管理数据库 `CMDB`)中提取配置参数并对配置模板进行渲染,从而产生设备配置命令集。最后,选择合适的连接管理插件

```
from netmiko import ConnectHandler
from nornir.core.configuration import Config
CONNECTION_NAME = "netmiko" # ④定义连接插件类标识符conn_name
class Netmiko: # ①重写连接插件接口open()方法
    def open(
        self,
        hostname,username,password,
        port,platform,extras,configuration
    ) -> None:
        parameters = {
            "host": hostname,"username": username,
            "password": password,"port": port,
            "ssh_config_file": configuration.ssh.config_file
        } # ②将网络连接赋值给连接插件类实例
        connection = ConnectHandler(**parameters)
        self.connection = connection
    def close(self) -> None: # ③重写连接插件接口close()方法
        self.connection.disconnect()
```

图 7 连接插件类实例

程序编写配置任务函数并在平台上执行。

应用于网络设备状态监视的一般流程见图 9。首先,基于 `TextFSM` 库编写设备状态查询指令相关的信息转换模板库。对于 `cisco`、`juniper` 及 `arista` 等厂商设备而言,开源项目 `ntc-templates` 以及预设开发相当规模的模板集,而对于华为和华三设备而言需要基于 `TextFSM` 重新开发。其次,对于 `Nornir` 返回的具体任务函数的返回结果采用相应信息转换模板进行处理,以此实现初级网络自动化过程中的结构化数据处理工作。

3.2 典型实例

为进一步验证该平台构建方法的正确性,采用华为模拟器 `eNSP` 创建实验拓扑环境,对其所属设备采用本文所设计的网络自动化运维平台进行批量配置和管理,结果表面其实际运行效果满足成规模设备自动化配置管理需求。囿于篇幅,整个实验环境及运行结果概况见图 10。

① 创建Jinja2配置模板	② 渲染Jinja2配置模板
<pre>{% for interface in interface_list %} interface {{ interface.name }} description {{interface.desc}} {% if interface.uplink %} port link-type trunk port trunk allow-pass vlan all {% else %} port link-type access port default vlan {{ interface.vlan }} {% endif %} {% endfor %}</pre>	<pre>interfaces=[{ "name": "GigabitEthernet0/0/1", "desc": "Uplink Port", "uplink": True }, { "name": "GigabitEthernet0/0/2", "desc": "Server port number one", "vlan": 100 }] config_interface=template.reader(interface_list=interfaces)</pre>
③ 利用平台运行配置任务函数	
<pre>from nornir import InitNornir from nornir_utils.plugins.functions import print_result from nornir_netmiko import netmiko_send_config def configure_switch(task): task.run(task=netmiko_send_config, config_commands=config_interface.split('\n')) nr = InitNornir(config_file="config.yaml") print_result(nr.run(task=configure_switch))</pre>	

图 8 设备配置通用流程

① 创建基于TextFSM的处理模板
<pre>Value INTERFACE (\S+) Value PHY (down !down up up\(\s\)) Value PROTOCOL (down \down up up\(\s\)) Value INUTI (\d*\.?\d*\%\ -) Value OUTUTI (\d*\.?\d*\%\ -) Value INERRORS (\d*) Value OUTERRORS (\d*) Start ^\\$*\${INTERFACE}\\$+\${PHY}\\$+\${PROTOCOL}\\$+\${INUTI}\\$+\${INERRORS}\\$+\${OUTUTI}\\$+\${OUTERRORS} -> Record ^PHY:isPhysical ^(?:\w+ ^)\\$down: ^(\w+):\\$+\\$+ ^Inuti:Oututi: ^Interface\\$+PHY\\$+Protocol\\$+Inuti\\$+Oututi\\$+inErrors\\$+outErrors\\$+\$\$ ^\\$*\$\$ ^. -> Error</pre>
② 利用平台运行监视任务函数
<pre>from nornir import InitNornir from nornir_netmiko import netmiko_send_command from nornir_utils.plugins.functions import print_result nr = InitNornir(config_file="./config/config.yaml") results=nr.run(netmiko_send_command,use_timing=10,use_textfsm=True,command_string='disp inter br') print_result(results)</pre>

图 9 设备状态监视通用流程

① 创建基于eNSP的实验拓扑
② 设备信息查询结果
<pre>netmiko_send_command***** * sw1 ** changed : False **** vvvv netmiko_send_command ** changed : False vvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvv INFO Quidway S5700-28C-HI Routing Switch uptime is 0 week, 0 day, 19 hours, 3 minutes ^^^^ END netmiko_send_command ^^^^^^^^^^^^^^^^^^ * sw2 ** changed : False **** vvvv netmiko_send_command ** changed : False vvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvv INFO Quidway S5700-28C-HI Routing Switch uptime is 0 week, 0 day, 18 hours, 51 minutes ^^^^ END netmiko_send_command ^^^^^^^^^^^^^^</pre>

图 10 平台运行实例

(下转第 345 页)

一种基于 Paramiko 的并行网络自动化编程方法

熊宗炬，李朝阳，李敬歌，张万彬，张小锋

中国西昌卫星发射中心，四川 西昌 615000

摘要：在对网络自动化运维技术的背景、概念、现状及相关支撑技术原理进行深入分析的基础上，提出一种面向传统网络运维的基于 Paramiko 的并行网络自动化编程方法，并对该方法中的核心模块、任务函数及接口函数的实现过程进行了重点阐述。应用实践表明，该方法通过融合 Python 生态中的多种开源技术，以快速高效完成网络操作为目标，在网络设备自动化批量操作与规模性能数据查询方面优势明显，具有良好的应用前景及推广价值。

关键词：Paramiko；网络自动化编程；TextFSM；并行处理；Jinja2

文献标志码：A

中图分类号：TP311.11

Parallel Network Automation Programming Method Based on Paramiko

XIONG Zongju, LI Zhaoyang, LI Jingge, ZHANG Wanbin, ZHANG Xiaofeng

Xichang Satellite Launch Center, Xichang, Sichuan 615000, China

Abstract: On the basis of in-depth analysis of the background, concept, current situation and related supporting technology principles of network automation operation and maintenance technology, this paper proposes a parallel network automation programming method based on Paramiko for traditional network operation and maintenance, and focuses on the implementation process of the core modules, task functions and interface functions in this method. The application practice shows that this method, by integrating a variety of open source technologies in the python ecosystem, aims to complete network operations quickly and efficiently, and has good application prospects and promotion value in automated batch operations of network devices and large-scale performance data query.

Key words: Paramiko; network automation programming; TextFSM; parallel processing; Jinja2

SDN (software definition network) 技术的发展和演进使得网络自动化编程技术在网络运维领域中发挥的作用愈发明显且重要，但实际生产网络中往往面临着设备不支持及自动化运维手段匮乏等困境。另据 Gartner 新近的一份技术报告提供的数据显示^[1]：截止 2021 年底，有 29% 企业或者组织从来不使用网络自动化，25% 的组织采用基于 CLI (command line interface) 的网络自动化（即初级网络自动化），13% 的组织仅在开发测试及实验环境中使用网络自动化，21% 的组织部分使用网络自动化，12% 的组织完全使用网络自动化。故此，

本文提出一种在传统网络与 SDN 共存时代面向初级网络自动化领域着眼解决企业高效运维手段匮乏及运维效率低下等难题的一种并行网络自动化可编程方法。

1 基础技术

1.1 初级网络自动化

通过 SSH 协议远程连接到不支持 NetConf 协议及 RESTConf 协议的网络设备而实现批量化操作管理的过程称之为初级网络自动化，属于基于设备级的网络开放可编程范畴。虽然初级网络自动化

作者简介：熊宗炬(1973—)，男，硕士研究生，正高级工程师，研究方向为数据网络通信、信息安全，E-mail: 5293826@qq.com；
李朝阳(1981—)，男，硕士研究生，工程师，研究方向为计算机软件。

与高级网络自动化相比在运维效率与实现容易度等方面存在不小差距,但考虑到传统网络架构目前仍然是企业生产网络应用与建设的主流,因此对于设备级初级网络自动化编程方法的讨论研究现实而重要。在 Python 开源生态领域,初级网络自动化工具包括开源开发库及开源自动化运维平台两类。典型的开源平台包括 Ansible、Salt 及 StackStorm;典型的开源库包括 Paramiko、Netmiko 及 Scrapy 等。综合考量设备适配性、运维灵活性与开发成熟度等因素,通常采用基于 Paramiko 进行设备级的初级网络自动化编程实践已然成为业界共识。

1.2 解析引擎 TextFSM

由于初级网络自动化方法基于 CLI 方式管理网络设备,其痛点在于所返回数据为非结构化形式,不利用快速计算与处理,而网络自动化技术发展的基础需求是要求设备尽量提供类似 JSON 格式的结构化数据。开源库 TextFSM 在网络自动化领域的应用恰好弥补初级网络自动化在实际应用中的缺陷,实现非结构化数据到结构化数据的有效转换。按照 TextFSM 官方的表述^[2],TextFSM 是一个采用 python 实现并基于状态机原理完成半格式化文本有效解析的模板引擎,其最初设计目标在于通过编程方式实现网络设备 CLI 命令行方式获取数据的结构化转换,当前已发展成面向此类数据结构化改造过程中的事实标准规范。该引擎以模板文件及解析对象文件作为输入经匹配操作后向用户提供 python 嵌套列表形式的结构化数据,其中模板文件与解析对象文件相互对应,用户可通过开发模板文件池方式而实现不同类别半格式化文件的批量解析操作。

1.3 Python 线程池

当 Python 程序需要创建大量生存周期短暂的

线程时,传统线程囿于操作系统交互成本而严重影响整个程序的执行效率。因此,采用 Python 线程池技术可解决上述困境。线程池在程序启动时便创建大量空闲线程,一旦向其提交任务函数,线程池则启动一个空闲线程执行该任务函数,执行完毕后该线程并不消亡,而是再次返回线程池成为空闲线程,等待执行下一任务函数。同时,线程池可有效控制并发线程数量,防止程序性能急剧下降或 Python 解释器崩溃^[3]。

Python 线程池的基类为 concurrent.futures.Executor,该类提供 ThreadPoolExecutor 及 ProcessPoolExecutor 两个子类,其中 ThreadPoolExecutor 用于创建线程池,而 ProcessPoolExecutor 用于创建进程池,二者为多核 CPU 硬件环境下的 Python 并行程序开发提供支撑。在实际应用中,线程池对象通过 submit 方法向线程池中添加相应任务函数,待任务函数执行完毕,submit 方法返回一个用于获取线程任务函数执行结果的 Future 对象实例。此外,Python 线程池实现了上下文管理协议,程序可通过 with 语句来管理线程池,从而避免手动关闭线程池而造成各种困扰。

2 结构设计

本文所设计的基于 Paramiko^[4]的并行网络自动化编程方法总体结构见图 1。整个结构设计包括核心模块、任务函数、接口函数、解析模板库及配置模板库等十部分。以高效设备运维为主线,本文提出的网络自动化编程方法在结构设计上可分为任务功能层面与辅助支撑层面。

功能层面主要任务概况为如何实现批量网络设备的快速高效操作管理。核心模块由 ExecCmdParamiko 及 ThreadRunner 两个功能类构成,Exce-

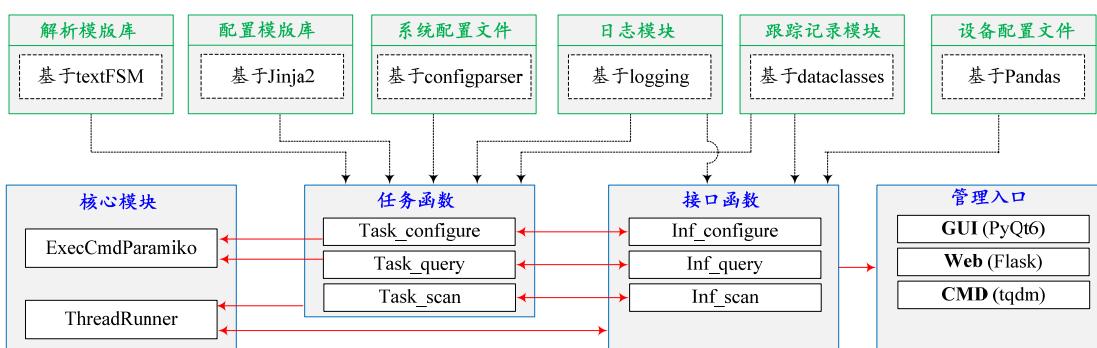


图 1 体系结构设计示意图

CmdParamiko 类是对开源库 Paramiko 的二次封装并向外提供更加简洁轻便的设备操作接口，ThreadRunner 是对 python 线程池的二次封装并为运维任务的并行处理提供支撑。任务函数按照应用类别分为配置类型 Task_configure、查询类型 Task_query 以及扫描类型 Task_scan 三类，其中扫描类型 Task_scan 任务函数是基于类 ThreadRunner 实现的批量并行 ping 操作用于设备在线情况探测。接口函数在类型上与任务函数一一对应，主要承担面向一组设备的任务函数规划调度、对线程池 ThreadRunner 运行结果的分析处理以及向管理接口输送任务函数运行结果三个主要职能。管理入口从类型上分为 GUI 类型、Web 类型及 CMD 类型，是管理用户的操作主界面。

辅助支撑层面主要任务概括为向任务函数及接口函数提供基础服务。其中解析模板库基于开源库 TextFSM 开发从而实现初级网络自动化领域内非结构化数据到结构化数据的转换。虽然业界已经存在面向 cisco 及 juniper 等设备的 ntc-templates 模板库，但其对华为及华三设备的支持度不高，故需直接基于 TextFSM 进行文本解析模板的开发。配置模板库基于 Jinja2 开发并用于网络设备的批量配置操作。系统配置文件主要包括 SSH 连接参数及设备登录参数等，通常寄存在文本文件中并通过 ConfigParser 工具类进行解析提取。日志模块基于 logging 模块设计并用于记录任务函数及接口函数在多线程运行模式下的内部事件。跟踪记录模块基于 python3.7 的新特性 dataclasses 实现，主要用于承载任务函数、接口函数及核心模块函数运行过程中产生的各类数据信息，是整个网络自动化编程方法设计中关于数据分析处理的基石。设备配置文件使用 excel 文件存储并基于 pandas 库^[5]进行设

备列表或设备分组的提取及过滤。

3 实现过程

3.1 辅助支撑模块

本文对第 2 章所设计的一种基于 Paramiko 的并行网络自动化编程方法在 Python 生态中进行编程实现，具体实现过程主要包括任务功能层面及辅助支撑层面。囿于篇幅，辅助决策层面重点阐述与任务函数编程关系紧密的跟踪记录模块的实现过程；任务功能层面将作为文章的重点予以阐述。

基于 python3 新特性 dataclasses 设计实现的跟踪记录模块各类之间内在关系^[6]见图 2。其中 Result 类是一个基础类并由此而衍生出 ResultSSH 类，ResultHost 类，ResultExecutor 类及 AggregateResult 类，ResultInterface 类用于对复合字典跟踪类 AggregateResult 类的遍历分析过程中，更多细节见表 1 所示。

3.2 任务功能模块

3.2.1 核心模块实现

核心模块包括类 ExecCmdParamiko 与类 ThreadRunner，旨在为设备 SSH 远控制、任务函数批量并行执行及任何函数运行结果反馈提供支撑工具，下文分别简述其编程实现要点。

(1) ExecCmdParamiko 类实现

ExecCmdParamiko 类是对基于 Python 实现 SSHv2 协议的开源库 Paramiko 的二次封装，旨在为远程控制网络设备操作提供更加便捷高效的编程服务，该类是配置类型与查询类型任务函数访问网络设备的入口。ExecCmdParamiko 类^[7]其内部结构见图 3，包括构造方法、设备登录方法、设备退

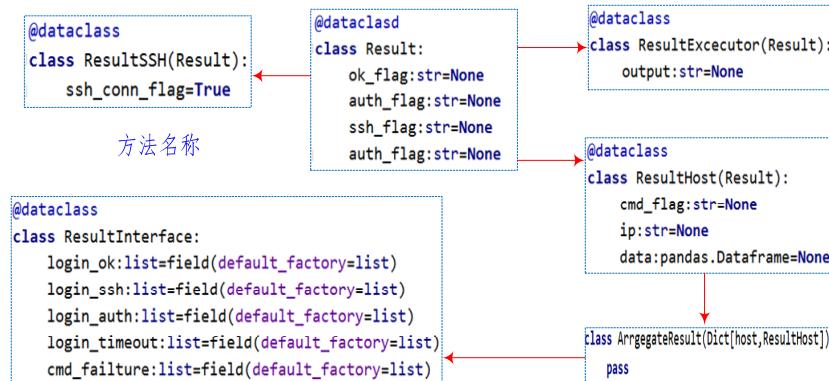


图 2 跟踪记录模块

表 1 跟踪记录模块类结构设计说明

类名	主要作用
Result	用于记录采用 SSH 方式远程登录网络设备可能发生各类正常及异常信息的基础类
ResultSSH	用于表征核心类库 ExecCmdParamiko 类中设备登录及退出方法的返回结果
ResultExecutor	用于表征核心类库 ExecCmdParamiko 类中单条及多条命令执行方法的返回结果
ResultHost	用于刻画配置类任务函数及查询类任务函数的返回结果信息
AggregateResult	刻画 ThreadRunner 类 run 方法的返回结果, 记录某任务函数在一组设备上的执行结果
ResultInterface	用户装载特定接口函数返回给管理界面的有效处理信息

出方法、单条命令执行方法及多条命令执行方法。其中, 方法 login_device() 返回 ResultSSH 实例对象用于承载 SSH 认证及登录信息, 同时一旦针对某台网络设备的 login_device() 方法执行成功则完成

命令执行句柄 CLI 实例变量的初始化, CLI 是后续远程执行所有指令的前提条件。方法 exec_single_cmd() 和方法 exec_multi_cmds() 返回 ResultExecutor 实例对象, 该对象不仅包含 SSH 方式登录设备时获得的各种提示信息也包括远程 CLI 方式执行指令所获得的半格式化回显文本(TextFSM 模板引擎的解析对象)。方法 exec_multi_cmds() 并不是 exec_single_cmd() 方法的简单循环重复, 二者或各种应用或组合使用, 譬如在华为系列交换机上用户期望统计到设备接口状态、设备接口描述以及接口所属 vlan 的对应关系, 则需要组合使用上述两种方法并分别获取不同 TextFSM 模板引擎所需的原始解析文本对象。

(2) ThreadRunner 类实现

ThreadRunner 类是对 python3 线程池技术的二次封装^[8], 旨在为完成特定功能任务函数在并行环境下的批量操作提供基础支撑服务, 其 python3 实现细节见图 4。其实例变量 self.number_thread 用于表征线程池大小, 通常在硬件性能允许的前提下其值等于等待执行特定任务函数的设备数组大小。run()

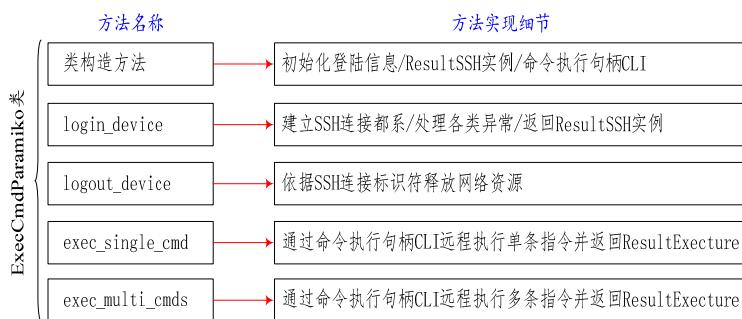


图 3 ExecCmdParamiko 类结构

```
def ThreadRunner(task, ipList):
    aggregateResult = {}
    tasks, results = [], []
    with concurrent.futures.ThreadPoolExecutor(max_workers=self.number_thread) as executor:
        for host in ipList:
            tasks.append(executor.submit(scan_host_inf, host))
        progress_bar = tqdm(concurrent.futures.as_completed(tasks), total=len(tasks), ncols=120)
        for each_task in progress_bar:
            try:
                each_ret = each_task.result()
            except Exception as e:
                print("执行task任务函数时异常:", e)
                results.append(ResultHost())
            else:
                results.append(each_ret)
                progress_bar.set_description("\n正在处理主机 {0} , 总体进度靠右--> ".format(each_ret.ip))
    for ret_obj in results:
        aggregateResult[ret_obj.ip] = ret_obj
    return aggregateResult
```

图 4 Python 实现的 ThreadRunner 类

方法为该类最重要的组件，其参数 task 代表任务函数实例，ipList 代表特定任务函数将在那些网络设备上被执行。ThreadRunner 类的实例化操作及其 run() 方法的调用在各类接口函数中完成，为使其在命令行模式下显示其管理操作的总体进度，可在 run() 方法中融入 tqdm 进度条工具呈现执行逻辑。

3.2.2 任务函数实现

通过网络编程方式在远程网络设备上完成特定配置或查询功能的一段代码集合称之为任务函数。通常情况，一个任务函数输入为一个网络设备管理地址而输出为 ResultHost 对象实例，相应设备 SSH 登录信息、文本回显示及数据分析处理结果都由 ResultHost 对象实例所承载。

通用任务函数实现流程见图 5，包括初始化、参数准备、处理分析与保存提交四个阶段。其中初始化阶段主要完成任务函数中调用的各辅助支撑模块的加载及初始化工作，譬如任务函数返回结果 ResultHost 的初始化、用于记录任务函数内部运行状态的日志记录器初始化、设备 SSH 方式登录信息的自动提取、相应 TextFSM 文本解析模板提取或配置 jinja2 模板的提取。参数准备阶段的目标在于完成网络设备配置指令集合的提取及创建 ExecCmdParamiko 实例对象(执行器)，一般情况下将需在网络设备上执行的指令以特定顺序存储在文本文件中。处理分析阶段主要针对 SSH 登录失

败情况，远程指令执行成功情况与指令执行失败情况分别进行处理分析。对于 SSH 登录失败或指令执行失败的情况，选择构造空载的 ResultHost.data 属性以此为后续接口函数在多线程结果处理中提供便利。对于具有查询功能的任务函数通常采用 pandas.DataFrame 类对其查询结果进行相应清洗处理并赋值于 ResultHost.data 属性。保存提交阶段主要针对退出网络设备前对相关资源进行释放并向调用任务函数的核心模块 ThreadRunner 返回 ResultHost 实例对象。

3.2.3 接口函数实现

接口函数在整个设计中起到内外衔接的作用，对内通过调用核心模块和特定任务函数实现在一组网络设备上以并行操作的方式实现批量指令操作，对外则向用户接口输出相应配置、查询及扫描操作的数据分析处理结果。通常情况下，一个接口函数输入为一组设备地址及某个特定任务函数对象实例，输出则为一个 ResultInterface 对象实例用于跟踪批量设备管理操作执行轨迹。

通用接口函数实现流程见图 6，包括初始化、并行操作及结果处理三个阶段。在初始化阶段主要完成接口函数返回结果 ResultInterface 实例对象的初始化、用于记录接口函数内部运行细节的日志记录器的初始化、特定任务函数加载以及网络设备参数提取。在本设计中将网络设备地址信息、分

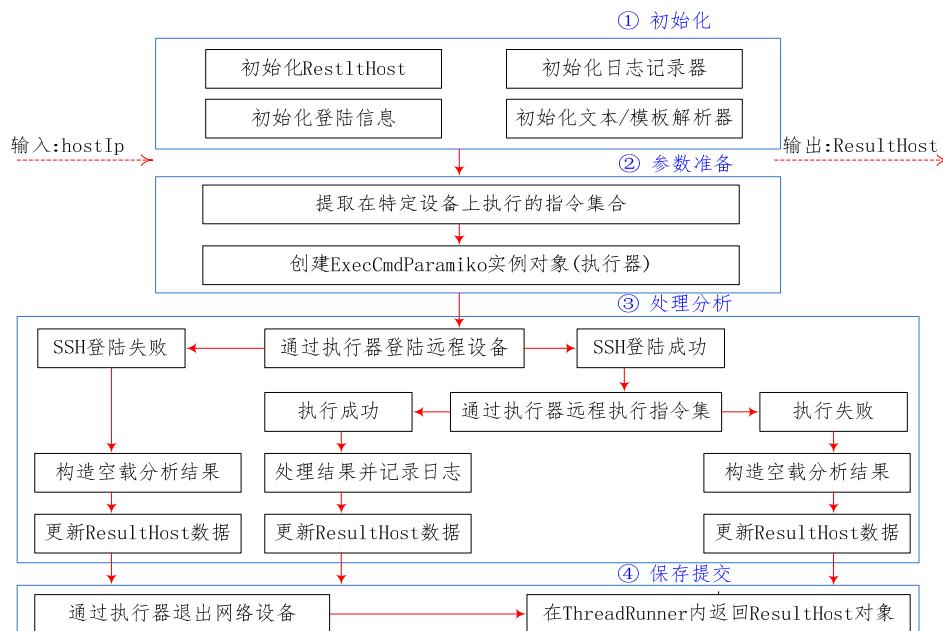


图 5 通用任务函数实现流程

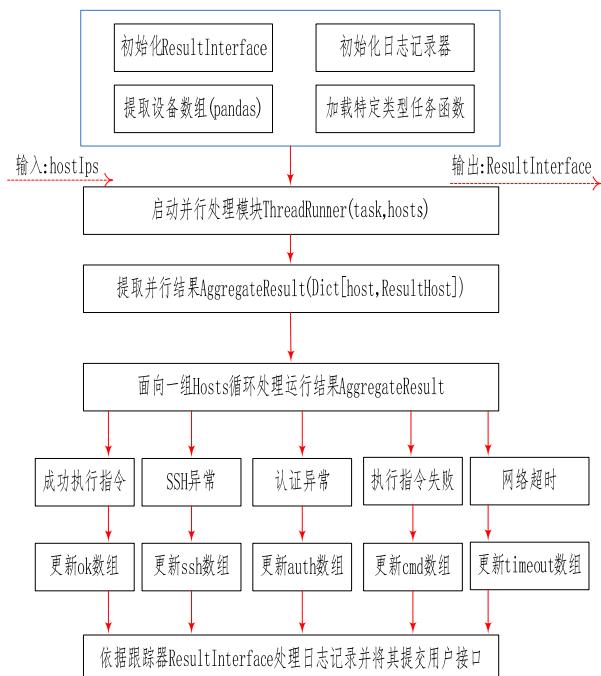


图 6 通用接口函数实现流程

组信息及辅助信息存储于 Excel 文件中并通过 pandas 进行特定设备数组的提取与过滤, 次方法相较于 YAML 或 XML 方式更为灵活实用。

并行操作阶段旨在完成两方面工作:(1)以设备数组及特定任务函数为参数, 通过调用核心模块中 ThreadRunner.run()方法实现特定任务函数在一组网络设备上的快速执行;(2)对多线程条件下的任何函数运行结果进行加工处理并最终构造形成字典型返回结果 AggregateResult(Dict[host, ResultHost]), 其键为网络设备管理地址信息, 对应值为特定任务函数在该设备上执行后的返回结果 ResultHost 实例对象。

结果处理阶段主要通过对 AggregateResult 实例对象的循环分析, 完成对多线程并行操作环境下特定任务函数在多个设备上执行情况的分类记录与处理, 从而为不同形式的用户管理入口提供数据展示支撑。上述五类情况包括设备执行成功情况、SSH 登录失败情况、设备远控认证失败情况、设备远控连接超时情况及远程指令执行异常情况。

4 方法应用

虽然 SDN 相关技术的商用阶段已在生产网络中大面积铺开, 尤其在数据中心网络领域。但鉴于很多企业或组织在技术先进性与资金投入等方面会做出权衡考量, 目前很多园区网或企业生产网并

没有对相关的网络基础设施进行升级换代, 因此本文所讨论的基于 Paramiko 技术的初级网络自动化在传统网络运维中的应用实践空间依然值得期待。

本文基于 Python 语言对所讨论的一种基于 Paramiko 的并行网络自动化编程方法进行了编程实现并应用于实际生产网络运维管理中。图 7 所示为采用本文所设计方法对 130 台交换机的端口状态的扫描探测, 以期快速找出每个端口的协议状态、物理连接状态、接口描述及所属 vlan 信息等。通常上述任务需在特定设备上连续执行多条指令并通过对比分析后方可获得结果, 可以看到采用本文所设计方法大概可在 50 s 左右完成所有任务。



图 7 方法应用实例

通过对核心模块中 ThreadRunner 类的实例变量 num_works 参数的设置, 可让基于 Paramiko 的网络自动化编程方法工作在串行模式与并行模式下。表 2 是对日常运维中的设备扫描、配置备份、接口统计、绑定用户信息统计以及网络设备软件版本信息查询等常用操作分别在串行模式与并行模式下相应任务函数执行的消耗时间, 其中执行操作的计算机配置是 CPU 为 Intel i5-3470, 内存为 8 GB, 系统为 Windows 10。综合而言, 运维效率提升明显。

表 2 测试环境说明

项目	设备数量	串行耗时/min	并行耗时/s
设备扫描	900	85	40
	2 500	262	55
配置备份	100	95	30
	150	155	32
接口统计	100	120	61
	150	185	62
绑定统计	100	113	50
	150	172	52
信息查询	100	98	65
	150	152	66

5 结束语

随着 SDN 技术给网络架构带来新的变革，关于网络自动化运维技术的应用与讨论亦日益广泛，但在传统网络运维领域，只有初级网络自动化技术及其工具才能发挥其作用及优势。本文正是基于上述事实并对相关自动化运维技术的概念、现状及支撑技术进行深入分析探讨，并提出一种面向传统网络运维的基于开源 Paramiko 工具的并行网络自动化编程方法。该方法通过融合 Python 生态中的多种开源技术，以快速高效完成网络操作为目标，在网络设备自动化批量操作与大规模性能数据查询方面优势明显，具有良好的应用前景及推广价值。

(上接第 215 页)

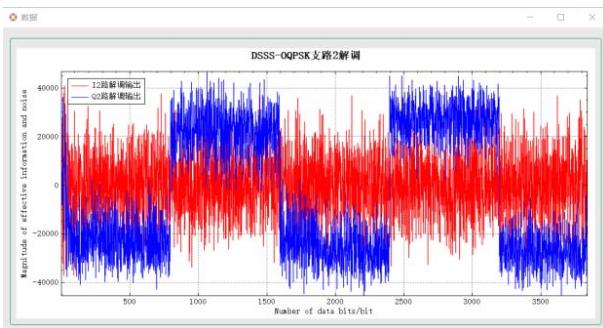


图 15 支路 Q 解调结果

比对可得二者一致，证明了信号解调结果的正确性。

5 结论

本文介绍了二代搜救信标信号的体制，主要包括信号的调制方式以及信号的帧结构，设计了软件模拟器的总体方案，介绍了软件的模块化设计方式以及其中主要功能模块的实现原理。利用二代信号软件接收机对生成的数字中频信号进行了测试与验证，对信号频谱以及捕获、跟踪和解调结果进行分析，解算得到的电文数据与设置的基带数据一致。验证结果表明模拟器软件基本实现了预期的功能。二代搜救信标信号模拟器的研究具有重要意义，能够为地面站接收机的研发过程提供可靠的测试数据与检测手段。目前利用模拟器产生的信号进行定位解算的结果精度还有待提高，将进一步对模拟器软件的算法进行优化和调整。

参考文献：

- [1] Inc.RedHat. Network automation for everyone[EB/OL]. (2022-02-02). <https://www.ansible.com/>.
- [2] Inc.Google. TextFSM documents[EB/OL]. (2022-03-05). <https://github.com/google/textfsm/wiki/TextFSM>.
- [3] Giancarlo Zaccone. Python parallel programming cookbook[M]. 2nd ed. Birmingham, UK: Packt Publishing Ltd, 2018: 35-50.
- [4] FORCIER J. Paramiko documentation[EB/OL]. (2022-03-05). <https://www.paramiko.org/>.
- [5] MCKINNEY W. Python for data analysis:data wrangling with pandas, Numpy, and IPython[M]. Beijing: China Machine Press, 2018: 115-135.
- [6] BARROSO D. Nornir tutorial documentation[EB/OL]. (2022-03-05). <https://nornir.readthedocs.io/en/latest/index.html>.
- [7] 王印. 网络工程师的 python之路[M]. Beijing: Publishing House of Electronics Industry, 2020: 130-155.
- [8] 肖明魁. Python 语言多进程与多线程设计探究[J]. 计算机光盘软件与应用, 2014, 17(15): 66-67.

参考文献：

- [1] 钱成, 李广侠, 赵陆文, 等. 北斗中轨卫星在海上搜救中的应用[J]. 卫星应用, 2015(5): 23-27.
- [2] 徐佳伟. 基于搜救信号的 TDOA/FDOA 定位技术研究 [D]. 南京: 南京邮电大学, 2022.
- [3] COSPAS-SARSAT. Cospas-Sarsat report on system status and operations No.34 (Jan-Dec 2017) [EB/OL].
- [4] COSPAS-SARSAT.Specification for second-generation Cospas-Sarsat 406-MHz distress beacons[EB/OL].
- [5] 郑晴, 王冰, 程宇新, 吴建军. 全球卫星搜救系统及其发展现状[J]. 卫星应用, 2015(10): 29-31.
- [6] 申翔, 吴培仁. 基于北斗卫星导航系统的海上应急搜救系统[J]. 指挥控制与仿真, 2018, 40(6): 43-49.
- [7] 中国卫星导航定位应用管理中心. 系统介绍[EB/OL]. <http://www.chinabeidou.gov.cn>.
- [8] 何浩, 李阳, 赵晨, 等. 搜救卫星系统技术发展与应用探讨[J]. 卫星应用, 2018(7): 26-31.
- [9] CAST Navigation LLC. CAST-GNSS specifications [EB/OL]. <https://castnav.com/gnssspecifications>, 2020.
- [10] Spirent Communications. Spirent GSS9000 series GNSS simulation system[EB/OL]. <https://www.spirent.cn/assets/u/datasheet-gss9000-series>.
- [11] 翟桅, 张国柱, 雍少为. 基本星座下北斗卫星导航系统服务性能分析[J]. 全球定位系统, 2011, 36(4): 56-60.
- [12] 齐巍, 常青, 张其善, 等. 高动态信号模拟器中的多普勒模拟算法[J]. 航空学报, 2008, 29(5): 1252-1257.
- [13] 张强, 赵齐乐, 章红平, 等. 北斗卫星导航系统 Klobuchar 模型精度评估[J]. 武汉大学学报(信息科学版), 2014, 39(2): 142-146.
- [14] 陈瑞琼, 刘娅, 李孝辉. 卫星导航系统中对流层改正模型分析[J]. 测绘通报, 2015(3): 12-15.
- [15] ZHANG Z T, HE X F, Applicability analysis of tropospheric delay correction model in rain and snow weather[J]. Journal of Navigation and Positioning, 2022, 10(2): 119-125.

基于 VXLAN 技术的城域网随遇接入网络架构研究

胡士斌，王晶，张琦，司长哲，冉兴隆

西昌卫星发射中心，四川 西昌 615000

摘要：一些大型企业城域网覆盖较大地域范围，承载多个业务系统和上千台入网终端设备，企业业务需要部分入网终端设备在不同网络节点入网，导致这些入网终端设备 IP 地址等入网信息需要经常更改调整，降低了工作效率，也存在修改错误的风险，网络随遇接入恰能解决相关问题。提出 VXLAN 技术结合 SDN、Overlay 等构建城域网网络架构，并应用 VXLAN 构建虚拟局域网，承载相应业务系统，为实现入网终端设备配置一个固定 IP 地址在不同入网节点随遇接入提供一个解决思路，同时也提高企业网络的灵活性和不同业务系统的安全性。

关键词：随遇接入；VXLAN；Overlay

文献标志码：A

中图分类号：TP391

Research on Architecture of On-Demand Access Network for Metropolitan Area Network Based on VXLAN Technology

HU Shibin, WANG Jing, ZHANG Qi, SI Changzhe, RAN Xinglong

Xichang Satellite Launch Center, Xichang, Sichuan 615000, China

Abstract: Some large enterprise metropolitan area networks cover a large geographical area, carrying multiple business systems and thousands of network access terminal devices. Enterprise business requires some network access terminal devices to be connected to different network nodes, resulting in frequent changes and adjustments to the IP addresses and other network access information of these terminal devices, which reduces work efficiency and also carries the risk of modification errors. Random network access can solve related problems. It proposes the use of VXLAN technologies combined with SDN, Overlay, etc. to construct a metropolitan area network architecture, and applies VXLAN to build a virtual local area network to host corresponding business systems. This provides a solution for configuring a fixed IP address for network terminal devices to access different network nodes as needed, while also improving the flexibility of enterprise networks and the security of different business systems.

Key words: on-demand access network; VXLAN; Overlay

一些大型企业覆盖地域范围大、设置业务节点多、部署业务系统复杂、网络安全要求高，同时还有数千台终端和设备入网，企业需要通过自建城域网来承载各业务系统数据传输和数千台入网终端设备接入。由于对网络安全有较高的要求，不能部署无线移动通信网络接入该城域网固网，各入网终

端和设备均设置静态 IP 地址。由于企业城域网早期构建，采用的传统网络设备和动态路由构建，使各个业务节点网络互联互通，该网络架构不同业务节点下用户网段是预先设置的，即各业务节点下入网终端 IP 地址须设置为预先分配的网段才能正常入网通信。

作者简介：胡士斌（1977—），男，高级工程师，主要从事网络通信工作；王晶（1994—），男，助理工程师，主要从事网络通信工作；张琦（1980—），男，工程师，主要从事网络通信工作；司长哲（1969—），男，高级工程师，主要从事网络通信工作；冉兴隆（1985—），男，工程师，主要从事网络通信工作。

在企业的现实业务中,部分入网终端和设备需要在多个业务节点移动部署入网,因此这些需要在多个节点入网的终端和设备预先分配多个 IP 地址,每个 IP 地址对应一个业务节点,在不同节点移动部署接入时,工作人员需对入网终端和设备的 IP 地址等入网信息进行重新设置。这种网络架构,不但造成 IP 地址管理困难,同时也因工作人员更改入网终端 IP 地址和网关等信息而对普通工作人员的网络素质提高了要求,经常重新设置 IP 地址的操作也容易提高误操作风险。普遍采用的移动 IP 技术解决方案^[1-2]也不适应该城域网特殊运用需求,因此在城域网中采用新的网络架构实现入网终端随遇接入功能是非常必要的,随着企业信息化建设发展,入网终端设备越来越多,对网络资源的管理、网络规模的扩展、网络的敏捷接入、业务的快速编排等有新需求。因此有必要研究新的网络架构,实现入网终端的随遇接入,以及企业信息化建设发展对网络通信新需求。

云计算已成为构建 IT 系统采用的新技术,服务器虚拟化是云计算关键技术之一,其在资源利用率、灵活性、扩展性等方面表现出强大优势。在服务器虚拟化技术应用的数据中心中,虚拟机需要在网络中自由迁移,对网络的灵活性提出了更高要求,传统的网络对虚拟机的应用有较大的限制作用,通过在传统物理网络上联合应用 Overlay 和 VXLAN 技术应运而生。目前, VXLAN 技术也逐渐在园区网中部署使用,主要解决当前传统园区网中灵活性较差的问题。

1 关键技术

1.1 Overlay 网络技术

Overlay 网络技术是一种在 Underlay 网络(基础网络)上叠加虚拟网的技术^[3-4],即在物理网络上叠加虚拟网络,不需关注 Underlay 网络采用的协议、技术,不需对 Underlay 网络进行大规模修改,只要能实现相关物理网络节点互联互通即能实现叠加虚拟网络的需求,实现业务应用在虚拟网络上承载。Overlay 网络技术叠加示意图如图 1 所示。

Overlay 网络具有独立的控制平面和转发平面,对于用户而言,物理网络是透明的,各业务系统之间可以实现逻辑隔离保护。Overlay 技术解决网络叠加的思路是将原始以太网报文承载到隧道

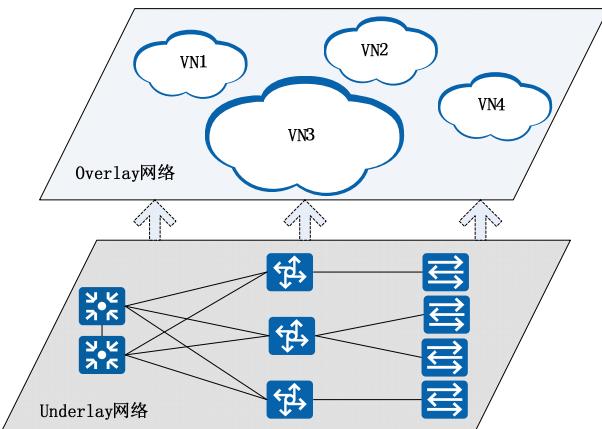


图 1 Overlay 网络技术叠加示意图

中传输。当前 Overlay 网络技术主要包括 VXLAN、GRE 等。Overlay 网络技术在网络扩展性、灵活性、多路径转发、安全性等方面都具有优势,同时通过网络叠加技术,可以实现对网络资源进行虚拟化,提高网络资源的灵活配置和高效利用。

1.2 VXLAN 技术

VXLAN(virtual extensible LAN, 虚拟局域网) 是一种网络虚拟化技术,目前在云数据中心中广泛应用。VXLAN 由 RFC7348 进行规范定义,采用 MAC-in-UDP 报文封装^[5-6],封装示意图如图 2 所示。VXLAN 通过隧道端点 VTEP (VXLAN tunnel endpoint) 对 VXLAN 报文进行封装和解封,对用户隐藏了隧道链路层帧数据的转发细节,其实现机制是,在发送端(隧道起始端)将终端/虚拟机发出的原始以太网报文封装在 UDP 中,然后再使用 Underlay 网络的 IP、MAC 地址作为外层头对原始以太网报文进行封装,在 IP 网络上传输。到达目的隧道终结点后,再由隧道终结点对报文进行解封,将解封后的原始以太网报文送达目标终端或虚拟机。

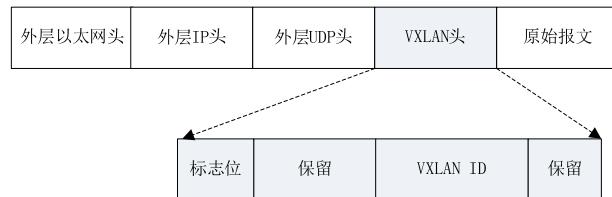


图 2 VXLAN 数据帧封装示意图

通过 VXLAN 技术的应用,入网终端、虚拟机不再需关注网络网络实现和链路层转发细节,虚拟网络可以提供更多的用户,同时用户也可以规划自

己的虚拟网络，提高了网络的灵活性，降低了网络管理的难度。

2 网络设计

2.1 总体架构

企业网络采用 SDN+Overlay 架构构建，网络架构如图 3 所示，分为转发层、控制层和应用层三层。

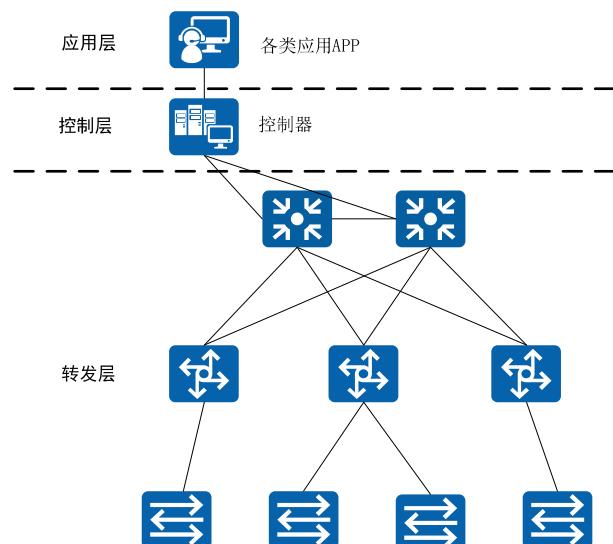


图 3 网络架构示意图

应用层：应用层主要用于调整、管理网络，业务编排部署等，应用程序可以由控制器本身提供、第三方提供，也可以是网络管理开发人员根据具体业务需求而开发的应用程序。

控制层：控制层主要指 SDN 网络控制器，也是 SDN 网络架构的核心，通过北向接口与应用层交互，通过南向接口与转发层交互，通过南向接口与转发层交互，能够获取整个网络的全局视图，将应用程序的需求转化为具体网络配置指令，控制转发层具体设备，从而实现对网络的集中管理和控制。

转发层：主要指具体的网络设备，根据控制层指令完成网络的调整和业务部署，实现数据报文的具体转发。Overlay 网络是在转发层基础网络上不同的虚拟网络，在各虚拟网络上承载不同的业务。

2.2 网关部署

网关部署有集中式网关和分布式网关。在集中式网关部署中，所有跨子网的网络流量均需要某一个三层网关设备转发，优点是便于网络管理和运

维，缺点是可能存在非最优转发路径，增加一些节点横向流量的路由交换时间，另外一个是网关设备性能要求较高。

分布式网关可分为下沉汇聚交换机网关和下沉接入交换机网关。跨子网的流量通过边缘设备（Edge）作为网关进行转发，便于扩展网络规模，且跨子网流量转发路径最优，但网关部署、故障定位及网络运维相对较为复杂。

2.3 VXLAN 及业务部署

VXLAN 隧道有静态隧道和动态隧道，静态隧道一般用于数量少、功能简单、网络规模小的网络中，一般在工程应用中均应用动态隧道，使用相应的控制面协议来实现隧道的建立和管理。VXLAN 控制面采用 BGP EVPN，不同 Edge 间需要建立 BGP EVPN 连接，以 Loopback 接口作为建立 BGP 连接和 VXLAN 隧道封装的源接口。

虚拟网络（VN）一般根据企业具体业务类型来划分，通过相应的网络应用程序来创建相应的虚拟网络实例。

子网划分与 VLAN 配置，在每个虚拟网络中可以根据需求划分多个子网，然后配置 VLAN 接入和 VXLAN 业务接入点，实现业务子网与虚拟网络的绑定接入。

3 网络模拟

网络模拟仿真主要是针对随遇接入功能的模拟仿真，随遇接入的模拟仿真主要是模拟设置同网段的两台 PC 机在不同网段节点交换机下可以互相通信。

网络模拟仿真平台采用华为 eNSP 1.3.00.100 对随遇接入功能进行简单模拟仿真。随遇接入功能模拟拓扑如图 4 所示。

3.1 Underlay 网络配置

在图 4 所示模拟拓扑中，DS02、DS03 作为 VTEP、VXLAN 的边缘设备，是 VXLAN 隧道的起点和终点。CS1 是透传设备，不对 VXLAN 报文做任何处理，只负责转发，Underlay 网络的配置主要是通过 IP 地址、路由配置实现 CS1、DS02、DS03 网络互通。按照表 1 配置表对 CS1、DS02、DS03 进行基本网络配置，实现 Underlay 网络互通。

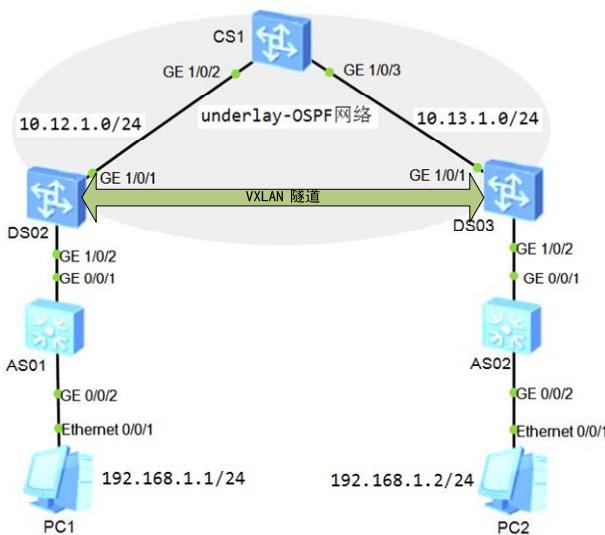


图 4 随遇接入功能模拟拓扑

表 1 Underlay 网络配置表

设备	端口	IP 地址	路由
CS1	LoopBack0	1.1.1.1/32	router id 1.1.1.1
	GE 1/0/2	10.12.1.1/24	network 1.1.1.1 0.0.0.0
	GE 1/0/3	10.13.1.1/24	network 10.12.1.0 0.0.0.255 network 10.13.1.0 0.0.0.255
DS02	LoopBack0	20.20.20.20/32	router id 20.20.20.20
	GE 1/0/1	10.12.1.2/24	network 20.20.20.20 0.0.0.0 network 10.12.1.0 0.0.0.255
	GE 1/0/2	10.13.1.2/24	network 30.30.30.30/32 network 30.30.30.30 0.0.0.0 network 10.12.1.0 0.0.0.255
DS03	LoopBack0	30.30.30.30/32	router id 30.30.30.30
	GE 1/0/1	10.13.1.2/24	network 30.30.30.30 0.0.0.0 network 10.12.1.0 0.0.0.255
	GE 1/0/2	10.13.1.1/24	network 10.13.1.0 0.0.0.255

3.2 Overlay 网络配置

Overlay 网络配置主要是在 Underlay 网络的基础上配置 VXLAN 隧道，建立起 DS02—DS03 之间的 VXLAN 隧道连接。由于图 5 所示拓扑较为简单，同时仅模拟一条 VXLAN 隧道，因此可以在模拟中采用静态方式建立 VXLAN 隧道。

在 VXLAN 隧道配置中，DS02、DS03 作为隧道端点，CS1 作为透传设备，仅需对 DS02 进行如表 2 配置。

```
<DS02>dis vxlan vni
Number of vxlan vni : 1
VNI          BD-ID      State
-----+-----+-----+
1           10         up
<DS02>dis vxlan tunnel
Number of vxlan tunnel : 1
Tunnel ID   Source      Destination      State  Type      Uptime
-----+-----+-----+-----+-----+-----+
4026531844  20.20.20.20    30.30.30.30    up    static    00:01:10
```

图 5 VXLAN 隧道建立情况

表 2 建立 VXLAN 隧道主要配置

设备	DS02	DS03
配置	Bridge-domain 10	Bridge-domain 10
	Vxlan vni 1	Vxlan vni 1
	Interface Nve1	Interface Nve1
	Source 20.20.20.20	Source 30.30.30.30
	Vni 1 head-end peer-list 30.30.30.30	Vni 1 head-end peer-list 20.20.20.20

3.3 业务接入

业务接入主要将业务终端的网络流量接入到相应的 VXLAN 隧道，主要是通过在 DS 创建子接口，绑定 BD 域，在 AS 中创建 Trunk 接口等，实现业务接入，具体配置如表 3 所示。

表 3 业务接入主要配置表

设备	PC1 接入	PC2 接入
PC	PC1: IP: 192.168.1.1/24	PC2: IP: 192.168.1.1/24
AS	AS01: vlan 10 interface GE0/0/1 port link-type trunk port trunk allow-pass vlan 10 interface GE0/0/2 port link-type access port default vlan 10	AS02: vlan 10 interface GE0/0/1 port link-type trunk port trunk allow-pass vlan 10 interface GE0/0/2 port link-type access port default vlan 10
DS	DS02 interface GE1/0/2.1 mode l2 encapsulation dot1q vid 10 bridge-domain 10	DS03 interface GE1/0/2.1 mode l2 encapsulation dot1q vid 10 bridge-domain 10

4 测试结果

测试主要针对 VXLAN 隧道建立以及在隧道两端 PC 互通情况。

隧道建立情况可以在 DS02 或 DS03 中通过命令查看，如图 5 所示。从图中可以看出，VN1 和隧道的状态均为 UP 状态，表明隧道建立正常。

在 PC1 命令行界面对 PC2 进行 ping 测试，测试结果如图 6 所示，从测试结果看，PC1 和 PC2

(下转第 366 页)

ARP 攻击原理及预警定位方法研究

张 琦，胡士斌，王 晶，何 攀，杨文凯

西昌卫星发射中心，四川 西昌 615000

摘要：目前 IP 网络承载着话音、图像等多种业务的综合传输，随着网络承载业务的增加，如何确保 IP 网络为数据传输提供合理、可靠的服务已成为必须深入研究的课题。在分析 ARP 原理的基础上，提出了 ARP 攻击预警与攻击源定位的方法，解决了如何在局域网中快速发现并定位 ARP 攻击的问题，确保了 IP 网络服务的安全稳定，更加有力地保障了任务中各种数据的传输。

关键词：ARP 攻击；ARP 协议；攻击探测；定位攻击

文献标志码：A

中图分类号：TP309.5

Research on ARP Attack Principle and Early Warning and Positioning Method

ZHANG Qi, HU Shibin, WANG Jing, HE Pan, YANG Wenkai

Xichang Satellite Launch Center, Xichang, Sichuan 615000, China

Abstract: At present IP network carries the comprehensive transmission of voice, image and other services. With the increase of network carrying services, how to ensure that IP network provides reasonable and reliable services for data transmission has become a topic that must be studied in depth. On the basis of the analysis of ARP principle, this paper puts forward the method of ARP attack early warning and attack source location, solves the problem of how to quickly find and locate ARP attack in the LAN, ensures the security and stability of IP network services, and more effectively guarantees the transmission of various data in the task.

Key words: ARP attack; ARP protocol; attack detection; positioning attack

ARP 协议是“address resolution protocol”（地址解析协议）的缩写，位于 OSI 网络七层模型中的数据链路层。ARP 协议是一个基础协议，是根据 IP 地址获取物理地址的一个 TCP/IP 协议。ARP 协议用于解析同一局域网或同一虚拟局域网内第 3 层协议地址和第 2 层硬件地址之间的映射关系^[1]。ARP 协议工作在数据链路层，由互联网工程任务组（IETF）在 1982 年 11 月发布的 RFC826 中描述制定，是 IPv4 中必不可少的协议，本身没有安全机制。正是由于协议本身不具备任何安全方面的机制，所以攻击者可以利用协议原理进行攻击。

文章将从 ARP 协议原理出发，分析 ARP 协议被攻击者利用的各种漏洞，并针对存在的问题进行分析并给出应对措施，达到从根本上解决 ARP 攻击预警及定位攻击源的目的。

1 ARP 攻击原理综述

1.1 ARP 协议原理

ARP 协议用于将 IP 地址解析为物理地址，当 IP 网内的数据要进行发送时，需要知道目标主机的物理地址，这时就需要通过 IP 地址找到相应的物理地址（MAC 地址）。

ARP 协议，即地址解析协议，是一种广泛应用于计算机网络中的协议，主要用于将 IP 地址与 MAC 地址进行映射，以实现网络设备之间的通信。以下是 ARP 协议的工作原理：

ARP 缓存表检查：当一台设备需要与另一台设备进行通信时，它首先会检查自己的 ARP 缓存表，看是否已经存储了目标设备的 IP 地址与 MAC 地址的映射。如果存在映射关系，则可以直接发送数据包给目标设备。

ARP 请求广播：如果 ARP 缓存表中不存在该

映射关系，那么设备就会发送一个 ARP 请求广播消息到局域网中的所有设备。这条消息包含了设备自身的 MAC 地址、目标设备的 IP 地址以及一个特殊的广播 MAC 地址。

ARP 响应：当其他设备接收到该广播消息时，就会将自己的 IP 地址与 MAC 地址回复给发送方，建立起新的 ARP 映射关系。发送方收到回复后，就可以将目标设备的 IP 地址与 MAC 地址存储到自己的 ARP 缓存表中，并使用它们进行后续的通信。

ARP 缓存表维护：ARP 协议通过动态更新 ARP 缓存表，实现了快速获取设备间地址映射的能力，提高了网络通信的效率。

综上所述，ARP 协议的工作原理是通过维护一个 ARP 缓存表，实现 IP 地址到 MAC 地址的映射。当需要通信时，设备会先检查 ARP 缓存表中是否有目标设备的 MAC 地址，如果没有，则通过广播发送 ARP 请求，收到请求的设备会发送 ARP 响应，从而建立起正确的地址映射关系。

1.2 ARP 协议原理缺陷

由于 ARP 协议是局域网协议，是建立在局域网内部各个终端设备相互可信的基础上，因此存在以下安全问题：

(1) 局域网内的各计算机终端通信是基于相互可信的，因而 ARP 应答包发送，无须经过任何认证。

(2) 计算机终端与网络传输设备中的 ARP 缓存表都是动态更新的，攻击者能够在 ARP 缓存表的老化时间内对缓存表中的信息进行恶意的修改，以实现假冒或其他恶意攻击。

(3) ARP 应答包没有确认或发送权限设置等安全机制，攻击者能够随意发送 ARP 应答包，当计算机机或网络传输设备接收到该应答包时将立刻刷新本地的 ARP 缓存表。

ARP 攻击利用的就是 ARP 协议不具备对 ARP 数据包进行安全确认的机制^[2]，攻击者可以随意发送虚假数据包对目标的 ARP 表项进行恶意修改。

2 ARP 欺骗攻击原理预警与定位

2.1 ARP 攻击示例分析

下面以图 1 为例，对 ARP 攻击流程进行分析，

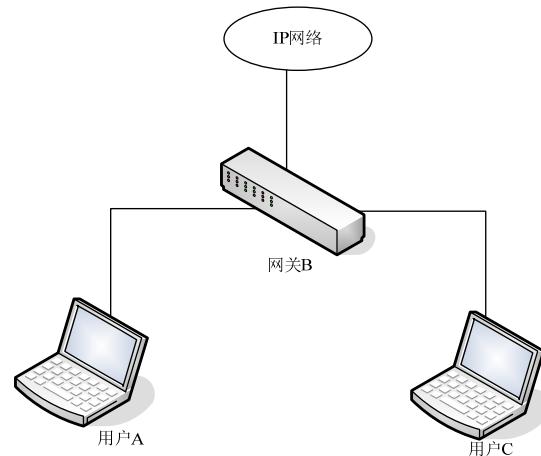


图 1 ARP 攻击流程分析

图中用户 A 为正常用户，用户 C 为攻击者，网关为交换机 B，攻击者 C 将通过 ARP 欺骗来控制用户 A 与网关 B 之间的通信。

此示例中 IP-MAC 地址用表 1 数据（为方便观看 MAC 地址为简单重复数字字母）。攻击流程如下：

(1) 攻击者 C 在 Windows 系统的命令提示符窗口中键入“ARP -a”，就可以获得这个攻击需要的基础信息：此网段主机和网关的 IP 和 MAC 地址。图 2 仅为命令示例。

(2) 欺骗用户 A

在知道了 ARP 协议存在的漏洞原理以及 ARP 数据请求与应答数据格式后，攻击者只需要编辑虚

表 1 IP-MAC 地址对照关系

计算机	IP 地址	MAC 地址
A	88.81.1.1	1A-1A-1A-1A-1A-1A
B	88.81.1.2	2B-2B-2B-2B-2B-2B
C	88.81.1.3	3C-3C-3C-3C-3C-3C

```
管理员: C:\Windows\system32\cmd.exe
Microsoft Windows [版本 6.1.7601]
版权所有 © 2009 Microsoft Corporation。保留所有权利。
C:\Users\Administrator>arp -a
接口: 192.168.0.10 --- 0xe
    Internet 地址          物理地址          类型
    192.168.0.1              2c-61-04-02-14-a8      动态
    192.168.0.9              7a-02-b2-ba-48-97      动态
    192.168.0.255            ff-ff-ff-ff-ff-ff      静态
    224.0.0.2                 01-00-5e-00-00-02      静态
    224.0.0.22                01-00-5e-00-00-16      静态
    224.0.0.251               01-00-5e-00-00-fb      静态
    224.0.0.252               01-00-5e-00-00-fc      静态
    239.192.152.143          01-00-5e-40-98-8f      静态
    239.255.255.250          01-00-5e-7f-ff-fa      静态
    255.255.255.255          ff-ff-ff-ff-ff-ff      静态
```

图 2 ARP 命令使用示例

假的 ARP 应答数据帧，控制计算机网卡进行应答数据帧的收发。在 windows 系统模式下，ARP 协议随时都会响应接收 ARP 应答。当计算机接收到 ARP 应答数据的时候，就会对本地的 ARP 缓存表立即进行更新，将应答数据中的 IP 和 MAC 地址存储在 ARP 缓存表中。因此，在示例的网络环境中，C 向 A 发送一个自己伪造的 ARP 应答，而这个应答中的数据为发送方 IP 地址是 88.81.1.2 (B 的 IP 地址)，MAC 地址是 3C-3C-3C-3C-3C-3C (B 的 MAC 地址本来应该是 2B-2B-2B-2B-2B-2B，这里被 C 的 MAC 地址伪造了)。当 A 接收到 C 伪造的 ARP 应答后，由于 A 并不知道地址是被 C 伪造的，就会更新自己的 ARP 缓存表项。现在数据传输设备 88.81.1.2 在用户 A 上的 ARP 映射信息被改变成了攻击源 C 的 MAC 地址，现在用户 A 发往数据传输设备的数据就会直接发给攻击者 C。

(3) 网关欺骗

对交换机的 MAC 欺骗：交换机与终端一样维护着一个动态的端口与 MAC 对应关系的缓存。交换机的工作原理是：首先，交换机内部有一个端口对应 MAC 列表，交换机的端口对应 MAC 列表记录着每一个端口下面存在哪些 MAC 地址，这个表初始状态是空的，交换机从各个端口的收发数据帧中学习。原理示例如下，当端口 1 所接的计算机发出了一条数据帧，这帧数据从端口 1 进入交换机，交换机就取这个数据帧的原 MAC 地址 1D-1D-1D-1D-1D-1D，然后在地址表中记录：端口 1 下面有 1D-1D-1D-1D-1D-1D。当端口与 MAC 对应表项建立后，所有发向 MAC 地址为 1D-1D-1D-1D-1D-1D 的数据帧，就全发给端口 1 的相连设备，而不会从其他的端口输出。与对计算机的 ARP 攻击相类似，如果把交换机上的端口对应 MAC 列表修改了，那么发给该 MAC 的数据就会发送给错误端口，导致用户数据丢失或发送给攻击者。例如，攻击者发送的伪造 ARP 应答数据包，发送者 IP 地址为原来用户 A 的 IP 地址，发送者 MAC 地址为 A 的 MAC 地址，交换机收到数据后，就会修改 ARP 缓存表把 A 的 MAC 与攻击者的端口相对应，当有上层数据需要向用户 A 传输时，交换机首先查 ARP 映射表，发现 A 的 IP 与和 MAC 与攻击者端口相对应，就会把发送给 A 的数据向端口 C 传输。这样，攻

击者 C 就能够获得发给正常用户 A 的数据。

(4) 数据包转发

经过上面两步，攻击者 C 已经能够获取用户 A 与网关 B 的通信数据，现在攻击突破口已经出现，接下来就是攻击者 C 进行数据包的转发，如果攻击者 C 再将收到的数据分别转发给用户 A 和 B，那么用户 A 和 B 将可以正常通信，但是他们不知道数据已经被 C 窃取。在实现数据转发功能之后，C 还可以对 A 和 B 通信的数据进行篡改、挂毒等操作，以进一步实施破坏，这就是著名的中间人攻击流程。

通过上面的示例分析，可以看出，最关键的第二步和第三步就是攻击者 C 把自己的 MAC 伪装成网关 B 和用户 A 的 MAC，就是通过发布这种虚假的 ARP 更新信息，使攻击者成为了 A 与 B 之间通信的中间人。因为在 ARP 缓存中的表项数据都有老化时间。默认情况下 ARP 缓存的老化时间是 120 s，如果再没有持续的信息更新，ARP 表项会自动去除。所以，为了保持虚假的 ARP 表，C 需要一直连续不断地发送虚假的 ARP 数据包。在这种情况下，当 C 停止发送虚假 ARP 数据包时，超过老化时间后 A 与 B 被修改的 ARP 表也将恢复正常，这将使攻击难以被发现和追查。所以，防范 ARP 攻击的重点就是及时发现，这就要求必须有足够的预警手段，下面将论述攻击探测的原理。

2.2 攻击探测器

根据上述对 ARP 攻击示例的分析，ARP 攻击将产生非法的 ARP 更新信息，这些非法的 ARP 更新信息将具有以下特征：

- (1) 非法的 ARP 更新信息与真实信息不符；
- (2) 非法的 ARP 更新信息会出现一个 MAC 地址对应多个 IP 的情形；
- (3) ARP 攻击者为维护虚假的 ARP 表，需要不停地发送 ARP 响应包；
- (4) ARP 攻击一般会伴随 MAC 地址漂移的现象（此点仅供参考）。

针对 ARP 攻击产生的这些非法的 ARP 更新信息特征，可以使用 ARP 攻击探测器进行对 ARP 攻击的进行预警。ARP 攻击的探测器的原理是在探测器上存放一张局域网 ARP 对应关系表，记录有

（下转第 297 页）

GEO SAR 对运动目标时序 RD 粗精度定位方法

付嘉瑜，李元昊，陈志扬，刘飞峰

北京理工大学 信息与电子学院 雷达技术研究院，北京 100081

摘要：SAR 目标定位技术是雷达摄影学中的一个重要分支，在军事侦查、地图测绘、高价值目标定位方面具有重要作用。传统距离多普勒（range-Doppler, RD）定位方法利用目标回波的时延信息和多普勒信息实现定位，但大多仅限于地面静止目标，无法较好地解决运动目标由于多普勒中心频率的变化造成定位精度恶化等问题。地球同步轨道合成孔径雷达（GEO SAR）由于所处的轨道高度极高、运动速度较慢，因此具备对目标的长时间观测能力，提出了一种 GEO SAR 对运动目标的时序 RD 定位方法，基于 RD 定位的原理利用两帧及多帧的时序信息联合实现对运动目标的定位。推导了时序 RD 方法的物像定位模型，基于简单运动场景进行了成像仿真及定位参数的联合估计，验证了方法的有效性；由于 GEO SAR 的特殊高度，方位向定位精度受径向速度和分辨率影响严重，通过提升径向速度的估计精度和优化分辨率，最终可有望实现运动目标百米级的定位精度。

关键词：GEO SAR；距离-多普勒定位模型；运动目标定位；定位精度

文献标志码：A

中图分类号：TP391

Coarse-Precision Moving Target RD Localization Method of GEO SAR Using Time Sequences

FU Jiayu, LI Yuanhao, CHEN Zhiyang, LIU Feifeng

Radar Technology Research Institute, School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China

Abstract: SAR target localization technology is an important branch of radar photography and plays a significant role in military reconnaissance, surveying and mapping, and high-value target localization. The traditional range-Doppler (RD) location method realizes localization by using the time delay information and Doppler information of the target echo, but it is mostly limited to the stationary scene on the ground and cannot well solve the problem of deteriorated location accuracy caused by the change of the Doppler center frequency of moving target. Geosynchronous orbit synthetic aperture radar (GEO SAR) has a long-time observation capability for targets due to its extremely high altitude and slow velocity. This paper proposes a sequential RD localization method for moving targets based on GEO SAR, which jointly realizes the localization of moving targets based on the principle of RD location using the time sequential information of two and multiple frames. This paper deduces the object-image location model of the sequential RD method, and conducts imaging simulation and joint estimation of location parameters based on a simple motion scenario. Due to the extensive observation slant range of GEO SAR, the azimuth location accuracy is severely affected by the estimation accuracy of the

基金项目：国家自然基金青年科学基金（62101039）；国家重点研发计划（2022YFB3901601）。

作者简介：付嘉瑜（1999—），男，博士研究生，研究方向为分布式星载 SAR 信号处理；李元昊（1990—），男，博士，预聘副教授，研究方向为星载新体制雷达系统和信息处理技术；陈志扬（1995—），男，博士，预聘助理教授，研究方向为合成孔径雷达系统；刘飞峰（1983—），男，博士，长聘教授，研究方向为雷达系统与信号处理。

radial velocity and resolution, verifying the effectiveness of the method. It is expected to eventually achieve the location accuracy of moving targets at the hundred-meter level.

Key words: GEO SAR; R-D location model; moving targets localization; location accuracy

目标定位，即雷达摄影定位，是 SAR 成像测量的基础，也是雷达摄影学的重要分支。其基本原理是利用 SAR 系统成像后像素点坐标和雷达发射接收的姿态、几何等信息，根据建立的相关定位模型及解算方法，实现目标位置坐标的准确解算。SAR 目标定位在军事侦察、导弹制导、地图测绘、高价值目标定位、自然灾害勘探等军民领域均具有重要作用。

传统 SAR 目标定位方法主要分为三种：立体定位方法、干涉测量定位方法以及距离-多普勒（RD）定位方法^[1]。立体定位方法^[2]是利用两幅 SAR 图像中心时刻的天线相位中心与目标位置间的几何关系和图像间同名像素点的坐标计算出目标的三维真实坐标，通常可分为同侧立体定位和异侧立体定位。近些年来，关于 SAR 立体定位方法的研究主要集中于定位模型的改进^[3-6]与定向参数的高精度解算等方面^[7-9]。干涉测量定位方法旨在利用 SAR 图像进行干涉处理得到的相位信息，从某一区域回波中提取出想要的高度及位置信息，具有精度高的优势，然而 InSAR 技术通常需要较高相干性，因此对于卫星轨道重复周期及天线的干涉基线提出了更高的要求^[10]。

利用单幅图像的距离-多普勒信息，结合地球模型，分析三者与成像区域目标的几何关系，从而实现定位的思想最早由 Curlander 提出^[11-12]，随即得到了广泛利用和发展，目前已成为 SAR 图像定位方法中的一种主流方法。基于距离-多普勒的定位方法无需依赖于地面控制点的选择，不受相干性的限制，其理论框架与处理流程与光学摄影测量技术具有相通性，因此具备前两种方法所不具备的优势。RD 定位实质就是求解地球模型约束下的距离多普勒方程，因此近些年的研究主要集中在模型优化与误差修正等领域。杨杰、李德仁等人^[13]在 RD 模型的基础上，通过引入数字高程模型（DEM）使得在无地面控制点的情况下，使得 ERS-1 天津地区的图像定位精度提升至 50 m 以内。孙耀天等人^[14]基于机载 SAR 平台的特点，提出了一种基于初始点选择的 RD 定位方法，提升了

RD 方程迭代求解的效率。刘秀芳和陆静等人^[15-16]还研究了地形、轨道、时延测量等因素对定位精度的影响。

综上所述，传统的 RD 定位方法只利用了图像的距离-多普勒信息，且当前研究的应用大多针对地面静止目标，利用 RD 定位方法解决星载 SAR 对运动目标的定位成为亟待解决的关键问题。2023 年我国成功发射了世界第一颗地球同步轨道 SAR（GEO SAR）卫星 LT4-01 星^[17]，GEO SAR 由于其独特的轨道可以形成对目标的长时间观测，因此在对距离、多普勒二维信息充分利用的基础上，通过引入时间维的信息有望解决运动目标的准确定位问题。本文针对 GEO SAR 对运动目标的定位问题，提出了一种时序 RD 定位方法，通过推导时序 RD 物像定位模型，发现在传统 RD 定位的基础上，利用两帧及多帧图像的时序信息可实现对运动目标的定位，通过简单场景的仿真，验证了所提方法的有效性，最终对定位精度的影响因素进行了分析。

1 GEO SAR 时序 RD 定位原理

GEO SAR 对运动目标的定位模型如图 1 所示，其中卫星和目标位置及速度均采用矢量形式来描述。其中， s 表示 GEO SAR 的位置矢量， v_s 表示 GEO SAR 的速度矢量， T 表示运动目标的位置矢量， v_T 则表示运动目标的速度矢量。为了简单起见，在后续的过程中假设运动目标为匀速直线运动的简单形式。

RD 定位的基本原理是利用目标回波的距离-多普勒信息，对目标进行定位，因此首先要建立定位坐标系，如图 1 所示，其中距离向的单位矢量为 R ，方位向的单位矢量表示为 x 。

1.1 GEO SAR RD 定位模型

RD 模型根据回波的距离历史和多普勒历史，将目标的实际距离与其在 SAR 图像中的位置建立了严格对应的物像关系，以下首先对 GEO SAR 运动目标定位的 RD 模型展开初步推导。

首先，GEO SAR 到运动目标的距离条件方程

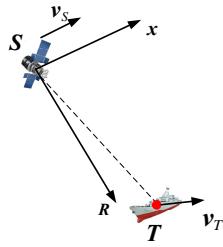


图1 GEO SAR 对运动目标定位示意图

如式(1)所示,其中 \mathbf{R}_{ta} 表示成像中心时刻目标 T 到天线相位中心 S 的斜距矢量,等式第二项表示成像中心时刻的斜距方程,等式第三项表示图像域中的距离向方程,其中 R_0 表示脉冲压缩的参考斜距, M_{range} 则表示距离向采样间隔, R_T 表示运动目标的距离向坐标。

$$\|\mathbf{R}_{ta}\| = \|\mathbf{S} - \mathbf{T}\| = R_0 + M_{range} \cdot R_T \quad (1)$$

雷达平台速度矢量和目标斜距矢量的乘积是多普勒频率的函数,因此可得到目标的多普勒频移条件方程为式(2)所示,其中 λ 为雷达波长, f_{dc} 为多普勒中心频率。

$$\mathbf{v}_S \mathbf{R}_{ta} = -\frac{\lambda R_{ta}}{2} f_{dc} \quad (2)$$

由于目标和雷达都处于运动过程,目标和雷达位置矢量需要进行高阶多项式拟合,高轨的位置矢量可根据精度需要展开到更高阶项的形式。由于目标匀速直线运动,因此只展开到一阶项即可。 t 表示方位时刻, \mathbf{S}_0 和 \mathbf{T}_0 表示初始时刻的GEO SAR和目标位置矢量, \mathbf{a}_S 表示GEO SAR的加速度矢量。

$$\begin{cases} \mathbf{S} = \mathbf{S}_0 + \mathbf{v}_S t + \mathbf{a}_S t^2 + \dots \\ \mathbf{T} = \mathbf{T}_0 + \mathbf{v}_T t \end{cases} \quad (3)$$

由于SAR的成像原理是对静止目标成像,因此需要在传统的成像模型基础上,考虑目标运动对信号斜距历史的影响,特别的,由于目标存在径向速度分量,将造成目标在图像域的位置偏移和方位模糊,如式(4)所示:

$$\begin{aligned} \exp\left\{-j\frac{4\pi R_{ta}}{\lambda}\right\} &= \exp\left\{-j\frac{4\pi}{\lambda} R_0\right\}, \\ \exp\left\{-j\frac{4\pi}{\lambda} \mathbf{n}_0 \mathbf{v}_S t_a\right\} \cdot \exp\left\{-j\frac{2\pi (\|\mathbf{v}_S\| t_a)^2}{\lambda R_0}\right\}, \\ \exp\left\{j\frac{4\pi}{\lambda} \mathbf{n}_0 \mathbf{v}_T t_a\right\} \cdot \exp\left\{-j\frac{2\pi (\|\mathbf{v}_T\| t_a)^2}{\lambda R_0}\right\}, \\ \exp\left\{j\frac{2\pi \|2\mathbf{v}_S \mathbf{v}_T\| t_a^2}{\lambda R_0}\right\} \end{aligned} \quad (4)$$

式中,单位矢量 \mathbf{n}_0 表示斜距方向,第二行和第三行的三项表示由于目标运动速度导致方位向信号引入相位项, $\exp\left\{j\frac{4\pi}{\lambda} \mathbf{n}_0 \mathbf{v}_T t_a\right\}$ 项会影响目标的多普勒中心频率大小,导致图像域沿方位向产生偏移, $\exp\left\{-j\frac{2\pi (\|\mathbf{v}_S\| t_a)^2}{\lambda R_0}\right\} \exp\left\{j\frac{2\pi \|2\mathbf{v}_S \mathbf{v}_T\| t_a^2}{\lambda R_0}\right\}$ 项表示由于目标运动速度产生的残余调频率,会导致图像在方位向发生散焦。

由于目标径向运动速度导致的方位向定位误差可以表示为 $\sigma_{x_{f_{dc}}}$,其大小可以表示为式(5)所示。其中, $v_T^r(t_a, \mathbf{v}_T)$ 表示目标运动速度在 t_a 方位时刻下径向速度大小,以远离径向的速度方向为正。

$$\sigma_{x_{f_{dc}}} = \frac{\lambda R_0}{2v_S} \sigma_{f_{dc}} = -\frac{R_0 v_T^r(t_a, \mathbf{v}_T)}{v_S} \quad (5)$$

综上所述,目标在定位坐标系中的真实定位坐标 (x_T, R_T) 与每帧图像的坐标 $(\tilde{x}_T, \tilde{R}_T)$ 可以建立物像模型,如式(6)所示:

$$(\tilde{x}_T^k, \tilde{R}_T^k) = (x_T^k, R_T^k) + \sigma_{x_{f_{dc}}}^k \cdot (\theta(x^k, x), \theta(x^k, R)) \quad (6)$$

其中, $\theta(\cdot, \cdot)$ 表示求第一个向量在第二个向量上的投影, x^k 表示第 k 帧序列的方位向方向矢量(零多普勒方向), x 表示之前定义的定位坐标系中的方位向坐标轴方向, R 表示定位坐标系的距离向坐标轴方向。 $\sigma_{x_{f_{dc}}}^k \cdot (\theta(x^k, x), \theta(x^k, R))$ 项表示将第 k 帧序列的偏移量投影到最终的定位坐标系方向,第 k 帧序列的偏移量作用在 x , R 两个维度。

1.2 时序RD粗定位方法

将多帧序列中所存在的时间维信息引入RD模型中,有助于对径向速度引起的位置偏移进行估计。因此,根据式(6)有时序RD粗定位模型如式(7)所示,其中在对GEO SAR的斜距进行近似时,采用了GEO SAR的匀速直线运动轨迹,忽略了加速度等高阶项对斜距的影响。

$$\begin{cases} \tilde{x}_T^k = x_T^k + \sigma_{x_{f_{dc}}}^k \cdot (x^k \cdot x) \\ \tilde{R}_T^k = R_T^k + \sigma_{x_{f_{dc}}}^k \cdot (x^k \cdot R) \\ R_T^k = R_T^1 + v_s^1 \cdot t_k - v_T \cdot t_k \\ v_s^k = v_s^1 \end{cases} \quad (7)$$

为了实现物像定位匹配,需要对定位模型中的位置参数进行估计,只考虑(x, R)二维平面的定位问题,则有 $(x_T^1, \tilde{v}_{Tx}, \tilde{v}_{TR})$,共计3个待估参数。在第一帧RD域,距离走动可通过包络对齐进行校正,因此有 $\tilde{R}_T^1 = R_T^1$ 。因此需要至少3个物像方程,才能对上述3个待估参数进行估计,列2帧物像方程实现定位如式(8)所示。为了使观测矩阵透明,进行相应的近似,令 $R_T^2 \approx R_T^1$, $\|\nabla f_d^2\| \approx \|\nabla f_d^1\|$, $t_2 - t_1 = \Delta t$,近似后可得矩阵形式的物像方程如式(8)所示,因此称为粗精度的定位方法。

$$\mathbf{L} = \begin{bmatrix} \tilde{x}_T^1 \\ \tilde{R}_T^1 \\ \tilde{x}_T^2 \\ \tilde{R}_T^2 \end{bmatrix} = \mathbf{H} \begin{bmatrix} x_T^1 \\ R_T^1 \\ \tilde{v}_{Tx} \\ \tilde{v}_{TR} \end{bmatrix} = \mathbf{H}\boldsymbol{\beta} \quad (8)$$

式(8)中, \mathbf{L} 为观测向量, \mathbf{H} 为观测矩阵, $\boldsymbol{\beta}$ 为待估向量,由于采用包络的定位方式受图像分辨率影响,因此需要考虑分辨率对定位误差的影响,采用加权最小二乘法进行估计,以提高估计精度,未知参数矢量的解如式(9)所示, \mathbf{C}_L 为图像定位误差的协方差矩阵,假设GEO SAR子孔径之间相互独立,非对角线上元素均为0,只体现对角线上的元素,因此可表示为式(10)。

$$\boldsymbol{\beta} = (\mathbf{H}^T \mathbf{C}_L^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{C}_L^{-1} \mathbf{L} \quad (9)$$

$$\mathbf{C}_L = \text{diag}(\sigma_{\tilde{x}_T^1}^2, \sigma_{\tilde{R}_T^1}^2, \sigma_{\tilde{x}_T^2}^2, \sigma_{\tilde{R}_T^2}^2) \quad (10)$$

最终,最小二乘估计的方差矢量可表示为式(11)所示:

$$\boldsymbol{\sigma}_{\boldsymbol{\beta}} = \text{diag}(\mathbf{H}^T \mathbf{C}_L^{-1} \mathbf{H})^{-1} \quad (11)$$

2 仿真验证

2.1 仿真参数设置

基于我国发射的第一颗GEO SAR卫星LT4-01星^[17]的相关轨道及系统参数,假设目标运动形式为匀速直线运动,开展成像仿真及定位研究,成像场景及部分场景参数如图2所示,其中具体的仿真参数如表1所示。

2.2 仿真结果

以GEO SAR对运动点目标进行正侧视成像为例,前后两帧图像的时间间隔 Δt 为60 s,成像结果

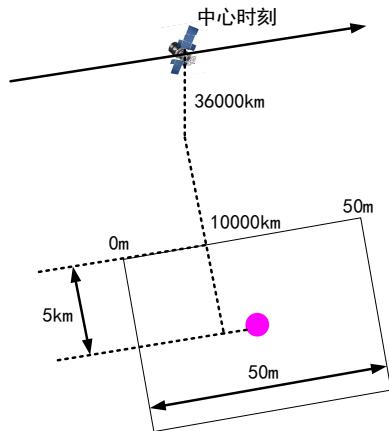


图2 GEO SAR与运动目标定位仿真示意图

表1 仿真参数

GEO SAR参数			
波长	0.24 m	速度	800 m/s
高度	36 000 km	合成孔径时间	60 s
脉冲重复频率	100 Hz	占空比	0.1
带宽	5 MHz	快时间采样率	15 MHz
前后两帧间隔时间	60 s	雷达运动轨迹	直线轨迹
目标运动参数			
目标初始位置	如图2所示	目标运动速度	[1,4] 地距向,方位向

如图3所示。第一帧图像如图3(a)所示,由于目标运动,方位向偏移量受中心时刻目标运动位置和径向速度的影响,已知目标地距向运动速度为1 m/s,则根据图2中的成像场景几何信息,计算得到目标径向速度大小约为0.2676 m/s,远离雷达方向,符号为负。根据式(5)计算可得方位向偏移量的理论值约为-12.4980 km,对应方位向的理论位置为-12.3780 km处。通过图像包络定位的方位向位置为-12.4556 km处,已知图像的方位向分辨率约为93 m,与理论位置偏差77.6 m在分辨率范围内,因此验证了方位向偏移量估计值与理论值的一致性,同理可验证第二帧图像如图3(b)所示。

在得到两帧图像目标位置的观测值后,代入到定位物像方程式(8)中,由于在正侧视观测下,距离向与方位向正交,导致观测方程形式进一步简化,如下所示:

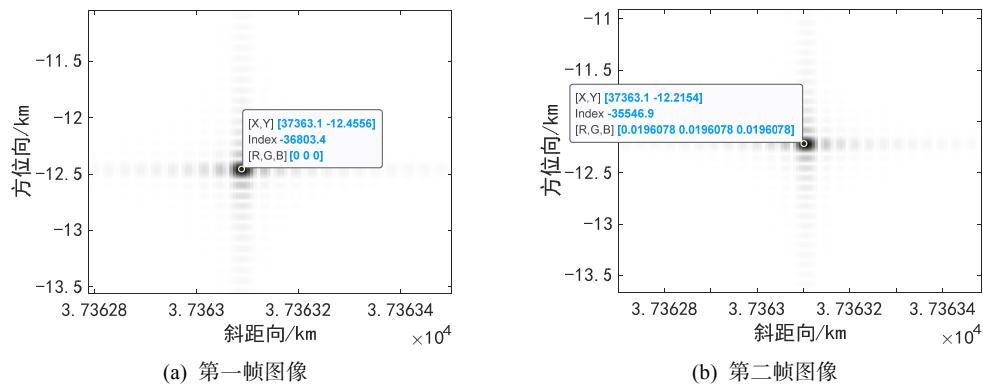


图 3 GEO SAR 对运动点目标定位成像序列

$$\begin{bmatrix} -12.455 & 6 \text{ km} \\ 37 363.088 \text{ km} \\ -12.215 & 4 \text{ km} \\ 37 363.11 \text{ km} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & -R_T^1/v_s^1 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 60 \text{ s} & -R_T^2/v_s^1 \\ 0 & 1 & 0 & 60 \text{ s} \end{bmatrix} \begin{bmatrix} x_T^1 \\ R_T^1 \\ \tilde{v}_{Tx} \\ \tilde{v}_{TR} \end{bmatrix} \quad (12)$$

除此之外，图像包络定位误差由分辨率决定，可表示为：

$$C_L = \text{diag}[(93 \text{ m})^2, (30 \text{ m})^2, (93 \text{ m})^2, (30 \text{ m})^2] \quad (13)$$

因此，代入式(9)可得真实位置估计值：

$$\beta = \begin{bmatrix} 3.14 \text{ km}, \\ 3.7363 \times 10^4 \text{ km}, \\ 4.0045 \text{ m/s}, \\ 0.3339 \text{ m/s} \end{bmatrix} \quad (14)$$

由式(11)可得估计的标准差向量为：

$$(\sigma_\beta)^{1/2} = \begin{bmatrix} 33.03 \text{ km}, \\ 30 \text{ m}, \\ 2.19 \text{ m/s}, \\ 0.707 \text{ m/s} \end{bmatrix} \quad (15)$$

从图 4 仿真结果可以看出，对运动目标的时序定位方法，方位向定位误差约为 3.02 km，距离向定位误差约为 30 m，方位向速度估计误差约为 0.0045 m/s，距离向速度估计误差约为 0.066 m/s。从时序物像方程的估计标准差矢量中可以看到，该方法对距离向速度的估计精度较差，从而导致方位

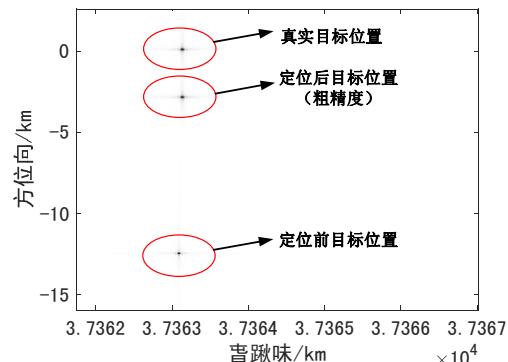


图 4 运动目标定位最终结果

向定位误差较大，往往不能满足高精度卫星定位的要求。

这种粗精度时序定位方法的综合定位及速度估计精度往往还受图像分辨率大小影响。针对不同径向速度估计精度对方位向定位精度的影响和不同图像分辨率对估计标准差向量的影响分析如图 5 所示，其中距离向和方位向分辨率不作区分。由于 GEO SAR 的长斜距影响，使得时序 RD 定位方法的估计精度对目标径向速度十分敏感。为了达到小于公里级的定位精度，需要径

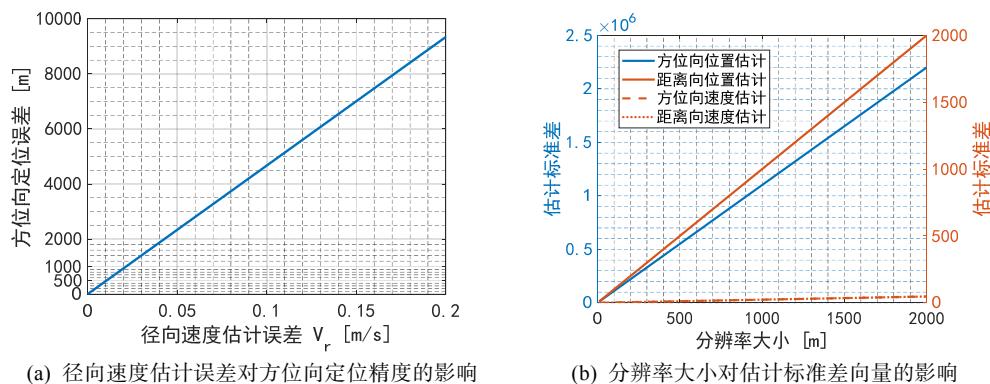


图 5 定位精度影响因素分析

向速度的估计误差小于 0.02 m/s，而在时序包络定位方法中，由于分辨率的限制，径向速度的定位精度通常大于 0.1 m/s，因此使得本定位方法的定位精度较差。

然而，基于时序 RD 的定位方法思路具有普适性和广泛性，为了进一步提高该方法的定位精度，该方法还可做如下改进：

(1) 利用时序多帧 RD 图像资源中的相位信息，采用子孔径沿轨干涉 (ATI) 或多普勒中心偏移估计 (DCA) 等方法，对目标的径向速度展开高精度估计。

(2) 上述模型只考虑了简单的运动形式，对于实际观测过程和目标运动过程中，应对式 (7)、(8) 中的物像关系模型进行高阶展开，考虑 GEO SAR 及目标运动过程中的非直线轨迹以及速度的高阶项，提高模型与实际运动过程的拟合精度。

3 结论

本文基于 GEO SAR 平台针对运动目标 RD 定位方法展开研究，利用 GEO SAR 可提供长时间观测的优势，在 RD 定位模型中利用距离、多普勒信息的基础上，引入时间维的信息，可实现对运动目标的定位及运动速度估计。

通过建立时序 RD 物像关系模型，推导了矢量形式的运动目标位置速度参数的估计模型，并利用最小二乘估计方法对待估参数进行了求解。以点目标为例，假设了一种简单的运动场景进行了仿真成像验证，在得到相隔 60 s 的前后两帧图像后，利用图像包络信息进行联合估计，最终得到目标方位向定位精度约 3 km，距离向定位精度约 30 m，方位向速度估计误差约为 0.0045 m/s，距离向速度估计误差约 0.066 m/s，方位向定位精度较差。由于方位向定位偏差主要由目标径向速度引起，其综合估计精度还与图像分辨率大小有关，因此对定位精度的影响因素展开了分析。通过分析可知，使得径向速度估计精度小于 0.02 m/s 时，该时序 RD 定位方法有望实现百米级的运动目标定位精度。

参考文献：

- [1] 胥遇时. 双(多)基地合成孔径雷达定位方法[D]. 成都:电子科技大学, 2018.
- [2] 张红敏. SAR 图像高精度定位技术研究[D]. 郑州: 解放军信息工程大学, 2013.
- [3] 黄国满, 赵争, 肖洲, 等. SAR 影像多项式正射纠正方法与实验[J]. 测绘科学, 2004(6): 27-30.
- [4] 尤红建, 丁赤飚, 付琨.SAR 图像对地定位的严密共线方程模型[J]. 测绘学报, 2007(2):158-162.
- [5] 李立钢, 尤红建, 彭海良, 等. 一种新的星载 SAR 图像定位方法的研究[J]. 电子与信息学报, 2007(6): 1441-1444.
- [6] 程春泉, 张继贤, 邓喀中, 等. 雷达影像几何构像距离-共面方程[J]. 遥感学报, 2012, 16(1): 38-49.
- [7] CHEN P H, DOWMAN I J. Space intersection from Ers-1 synthetic aperture radar images[J]. The Photogrammetric Record, 1996, 15(88): 561-573.
- [8] PU-HUAI C, DOWMAN I J. A weighted least squares solution for space intersection of spaceborne stereo SAR data[J]. IEEE Transactions on Geoscience and Remote Sensing, 2001, 39(2): 233-240.
- [9] 吴颖丹, 明洋. 缺少地面控制点的多源 SAR 影像联合定位方法[J]. 湖北工业大学学报, 2012, 27(4): 92-96.
- [10] 皮亦鸣, 杨建宇, 付毓生, 等. 合成孔径雷达成像原理[M]. 成都: 电子科技大学出版社, 2007.
- [11] CURLANDER J C. Location of spaceborne sar imagery[J]. IEEE Transactions on Geoscience and Remote Sensing, 1982, GE-20(3): 359-364.
- [12] CURLANDER J C. Utilization of spaceborne SAR data for mapping[J]. IEEE Transactions on Geoscience and Remote Sensing, 1984, GE-22(2):106-112.
- [13] 杨杰, 潘斌, 李德仁, 等. 无地面控制点的星载 SAR 影像直接对地定位研究[J]. 武汉大学学报(信息科学版), 2006(2): 144-147.
- [14] 张耀天, 孙进平, 毛士艺. 机载 SAR 图像直接定位算法研究[J]. 信号处理, 2009, 25(4): 669-673.
- [15] 刘秀芳, 刘佳音. 星载 SAR 图像的定位精度分析研究 [J]. 遥感学报, 2006(1): 76-81.
- [16] 陆静, 郭克成, 陆洪涛. 星载 SAR 图像距离-多普勒定位精度分析[J]. 雷达科学与技术, 2009, 7(2): 102-106.
- [17] JONES A. China launches first geosynchronous orbit radar satellite[EB/OL]. <https://spacenews.com/china-launches-first-geosynchronous-orbit-radar-satellite/>.

基于像素分辨率映射的单目人员三维定位

于振，张业欣，唐诗洋^{*}，冯杰

中国电力科学研究院 信息通信研究所，北京 102211

摘要：基于视觉图像提取人员三维位置信息对于智能安全管控系统是十分重要的，提出一种基于单目视觉的人员三维定位跟踪方法，通过对行走人员的位姿节点识别和基于图像分辨率的空间映射（PR-M），解决单目视觉缺少先验人员高度情况下的行人高度估计和人员定位问题。所提出的单目视觉的人员三维定位方法在不依赖于特殊标定样式的情况下，通过建立像素点真实大小和像素点在现实世界三维空间的 XY 方向分辨率及 Z 方向分辨率的映射关系，结合人员位姿关键节点进行空间信息的升维，实现人员的三维定位。该方法在实际变电站电力生产场景进行测试，测试结果表明该方法的人员高度估计准确率约 98%，30 m 内的平均人员定位误差约 17.4 cm，可以为变电站智能安全管控提供人员位置风险管控。

关键词：视觉定位；位姿识别；高度估计；图像分辨率映射

文献标志码：A

中图分类号：TP391

3D Monocular Personnel Positioning Based on Pixel-Resolution Mapping

YU Zhen, ZHANG Yexin, TANG Shiyang^{*}, FENG Jie

Department of Information and Telecommunication, China Electric Power Research Institute, Beijing 102211, China

Abstract: Extracting three-dimensional positional information of individuals based on visual images is crucial for intelligent security control systems. This paper proposes a monocular vision-based method for 3D positioning and tracking of individuals. By identifying pose keypoints of walking individuals and using a spatial transforming based on pixel-resolution mapping (PR-M), it addresses the challenges of pedestrian height estimation and individual positioning in monocular vision systems, particularly in the absence of prior knowledge about the individual's height. The proposed monocular vision-based 3D positioning method does not rely on special calibration patterns. Instead, it establishes a mapping relationship between the actual size of pixel points and the resolution of pixel points in the XYZ directions in the real-world 3D space. This, combined with the keypoints of the individual's pose, enables the dimensionality elevation of spatial information and achieves 3D positioning of individuals. This method is tested in a real substation power production environment. The test results show that the average positioning error within 30 meters is around 17.4 cm. Experimental results show that this method can provide risk control of personnel positioning for intelligent safety management in substations.

Key words: vison-based positioning; pose estimation; height estimation; pixel-resolution mapping

人员定位技术是智能化生产安全管理系统的
核心，在当前社会的许多行业发挥着重要作用。在
煤矿行业，文献[1]提出了一种基于双目视觉的矿
井人员定位，文献[2]提出了一种基于深度视觉的

矿井巷道人员定位方法；在电力生产行业，文献[3]
提出了一种水电厂入侵人员定位方法，文献[4]采
用超宽带定位技术（UWB）在火电厂开展定位实
验；在电网行业，文献[5]基于 UWB 技术开发了变

基金项目：北京市科技新星计划（20220484092）。

作者简介：于振（1981—），男，硕士研究生，高级职称，研究方向为电力安全与应急技术；张业欣（1997—），男，硕士研究生，研究方向为深度学习、电力安全技术。

电站可视化管控系统，文献[6]通过引入无迹卡尔曼滤波改进变电站 UWB 定位抖动问题，文献[7]通过鲸鱼算法优化变电站多目标定位执行过程，文献[8]提出了一种基于声信道分析的电缆隧道内人员定位技术；在建筑行业，文献[9]提出了一种 5G+蓝牙融合的人员定位技术，文献[10]将蓝牙定位和建筑信息建模结合，从而动态评估建筑工人轨迹安全。

位置信息是人员安全控制的核心要素之一。传统的室内定位方案主要分为蓝牙定位^[11]、无线传感网络^[12]和 UWB^[13]等定位方案。其中基于 UWB 的人员定位技术路线因其高精度的特点，已在许多行业落地相关应用。但受制于大量定位基站的部署和维护成本，加之相对繁琐的使用流程，使得该技术方案难以大范围推广应用。

以视觉传感器为基础的被动式定位方案由来已久，文献[14]于 1991 年通过调节相机焦距的方法确定相机与目标之间的距离关系以完成目标定位，文献[15]提出了一种灵活的相机内外参数的标定方法，为后续的视觉人员定位奠定基础。近年来，得益于深度学习的快速发展，基于视觉的定位技术的相关研究开始备受关注。文献[16]提出了一种基于集成学习的视觉定位方法，文献[17]提出一种适用于变电站环境的单目人员定位方法，文献[18]针对电力场景的人员遮挡问题提出了一种鲁棒的人员定位方案。其他视觉定位路线也被广泛探索，文献[19]提出了一种双目视觉定位方案，文献[20]使用多台 RGBD 相机识别别人身体关键节点完成人员定位，文献[21]通过融合音频和视频信息提升定位精度，文献[22]通过背景差分实现基于全景相机的空间感知，文献[23]通过人体轮廓和深度相机实现人员定

位，文献[24]提出采用视觉信息和 WiFi 信息融合的定位方法，文献[25-27]提出一种图像分辨率映射方法，通过对场景的三维重建实现目标定位。

然而对于实际生产环境下的安全管控而言，视觉定位方案不得不在精度和现场部署复杂度这两个要素下权衡。单目视觉技术方案往往比较简单，通常采用特殊的标定样式，但往往精度较低。双目或利用激光雷达、与先扫描好的场景三维模型等进行联合标定定位^[28-29]，虽然精度会略有提高，但就现实场景而言，大量的生产环境内视频监控摄像头彼此孤立构成了无数个单目画面，往往对现场的生产带来过大负担。并且以上方法通常需要根据先验的人员高度信息等去实现人员三维定位，这带来了更大的定位误差。

综上，本文提出了一种基于像素分辨率映射的人员定位方案，旨在：(1) 通过图像分辨率映射和人员位姿预测实现人员高度的准确估计，(2) 通过单目视觉在缺少人员高度等先验信息下实现一种高精度的人员三维实时定位方案。

1 定位模型

在单目视觉下实现零人员高度先验信息的高度是十分困难的，本文针对该问题提出了整体定位模型如图 1 所示。在单幅画面内，进行人员位姿估计识别其关键节点，并利用图像分辨率映射方式根据其所估计的关键节点信息来推测人员高度，最终利用相机内外参数结合关键节点像素实现单目人员三维定位。

1.1 位姿识别

本文为获取人员踝、膝、胸和头等关键节点的

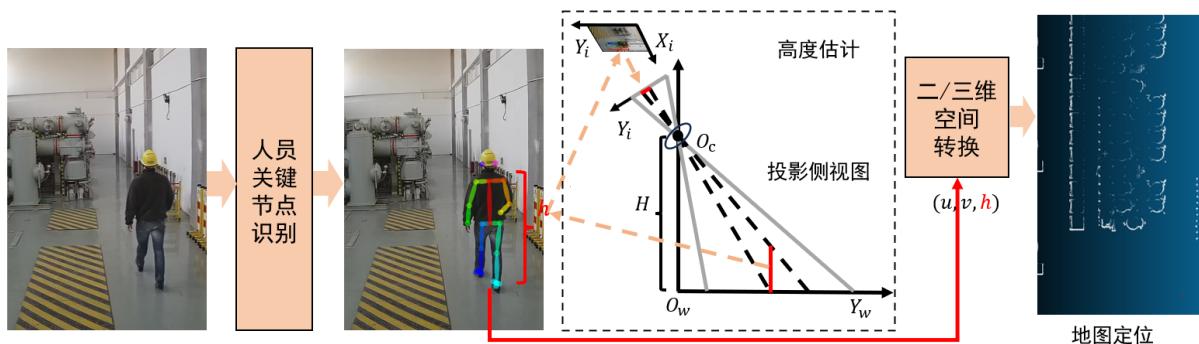


图 1 算法整体流程

二维坐标，采用 Openpose^[30]进行人员相关节点的位姿估计。Openpose 的网络结构首先通过 VGG19 网络的前 10 层，获取特征图 F ，作为后续阶段的输入。在第 1 阶段，有两个分支：分支 1 负责预测热图 S ，它指示关键点的位置；分支 2 负责预测向量图 L ，它表示关键点之间的关系。在第 1 阶段之后， S 和 L 与 F 合并，作为下一阶段的输入。后续阶段与第 1 阶段类似，通过重复这一过程可以提高热图的精度，并使肢体向量更加清晰。

Openpose 的另一个核心贡献是 PAF (part affinity fields)，它表示图像中每个像素的关联性，对于大小为 $W \times H$ 的图像，如果某个像素属于某个肢体，其 PAF 将被表示为指向该肢体方向的单位向量。在提供的公式中， k 代表第 k 个人， $L_{c,k}$ 代表第 k 个人的第 c 条肢体，而 p 代表当前像素，如公式 (1) 所示：

$$L_{c,k}^*(p) = \begin{cases} v, & \text{if } p \text{ on limb } c, k \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

在此之下：

$$v = \frac{(x_{j_2,k} - x_{j_1,k})}{\|x_{j_2,k} - x_{j_1,k}\|_2} \quad (2)$$

根据公式 (2)，如果点 p 属于 c 肢体，则点 p 的 PAF 受到公式 (3) 的约束。

$$0 \leq v \cdot (p - x_{j_1,k}) \leq l_{c,k}, |v \cdot (p - x_{j_1,k})| \leq \sigma_l \quad (3)$$

1.2 高度估计

1.2.1 相机模型

相机模型如图 2 所示，由图可见相机被放置场景内一定高度的位置，该高度即为相机镜头中心与世界空间内 $X_wO_wY_w$ 平面的垂直高度。对于一般场景设置而言，相机通常以一定角度照向地面，不妨设其俯仰角为 ϕ ，水平旋转角为 θ ，摆动角为 ψ 。为使问题略微简化，对于摄像头不妨设 $\psi=0^\circ$ 和成像器的中心位于光轴上。前者可以通过固定摄像头做到，也是通常安全监控摄像头的常见配置，后者则是目前主流摄像头的基本属性。

如图 2 所示，世界坐标系原点位于镜头光心于世

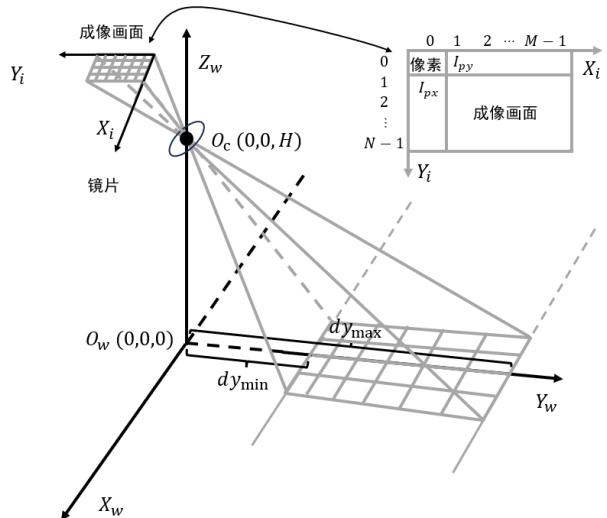


图 2 PRM 相机模型成像情况

界空间的 $X_wO_wY_w$ 平面上， Y_w 与光轴在 $X_wO_wY_w$ 上的投影重合，镜头光心 O_c 处于 $(0,0,H)$ ， H 为相机高度。焦距 f 和高度固定下， dy_{\min} 为成像画面可以成像的最小距离， dy_{\max} 为最大成像距离，二者之外的位置无法被映射。画面像素假设其在 X_iY_i 方向分别包含 $M \times N$ 个像素，像素的长宽为 I_{px} 和 I_{py} 。图 2 基本展示了画面像素与世界空间的映射关系的建立过程。

1.2.2 图像分辨率映射

如果每一个像素对应世界空间的位置可以确定，那么从二维画面提取人员或任意目标的高度、位置等信息都将可行，下面引出图像分辨率映射确定任意像素对应空间位置的映射。

像素分辨率与空间的映射分为两个方向，即 X_i 和 Y_i 方向，其像素分辨率分别表示为 R_x 和 R_y ，如图 3 所示。以 Y_i 方向下的分辨率 R_y 为例，其能表示的长度为 dy_{range} 如公式 (4) 所示：

$$dy_{\text{range}} = dy_{\max} - dy_{\min} = \left(\tan\left(\phi + \tan^{-1}\left(\frac{I_{py} \times N}{2f}\right)\right) - \tan\left(\phi - \tan^{-1}\left(\frac{I_{py} \times N}{2f}\right)\right) \right) \times H \quad (4)$$

如图 3， Y_i 方向第 n 个像素所对应的分辨率 $R_{y(n)}$ 表示如公式 (5) 所示：

$$R_{y(n)} = \tan \left[\phi - \tan^{-1} \left(\frac{I_{py} \times N}{2f} \right) + \sin^{-1} \left(\frac{(n+1) \times I_{py} \times \sin \left(\tan^{-1} \left(\frac{2f}{I_{py} \times N} \right) \right)}{\sqrt{\left(\frac{I_{py} \times N}{2} - (n+1) \times I_{py} \right)^2 + f^2}} \right) \right] \times H -$$

$$\tan \left[\phi - \tan^{-1} \left(\frac{I_{py} \times N}{2f} \right) + \sin^{-1} \left(\frac{n \times I_{py} \times \sin \left(\tan^{-1} \left(\frac{2f}{I_{py} \times N} \right) \right)}{\sqrt{\left(\frac{I_{py} \times N}{2} - n \times I_{py} \right)^2 + f^2}} \right) \right] \times H$$
(5)

类似的，对于固定 Y_i 方向上， X_i 方向某一行像素分辨率如公式 (6) 所示：

$$R_{x(n)} = \frac{\sqrt{1 + \left(\tan \left(\phi - \tan^{-1} \left(\frac{I_{py} \times N}{2f} \right) + \sin^{-1} \left(\frac{\left(n + \frac{1}{2} \right) \times I_{py} \times \sin \left(\tan^{-1} \left(\frac{2f}{I_{py} \times N} \right) \right)}{\sqrt{\left(\frac{I_{py} \times N}{2} - \left(n + \frac{1}{2} \right) \times I_{py} \right)^2 + f^2}} \right) \right)^2}} \times H \times I_{px}$$
(6)

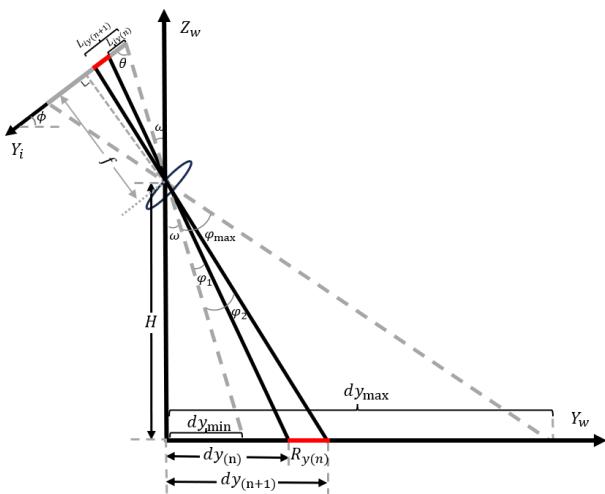


图 3 Y 方向分辨率映射投影侧视图

1.2.3 高度估计

仔细分析公式 (5)、(6) 可以发现，对于已知摄像头而言，即 $\phi, N, I_{px}, I_{py}, f, H$ 全部已知，假设人员站立，其与地面垂直，则分析其高度等同于在 Y_i 方向计算像素分辨率 R_y ，如图 4 所示，其计算方式如公式 (7)。

$$h_{\text{person}} = \frac{dy_{(\text{head})} - dy_{(\text{foot})}}{dy_{(\text{head})}} \times H$$
(7)

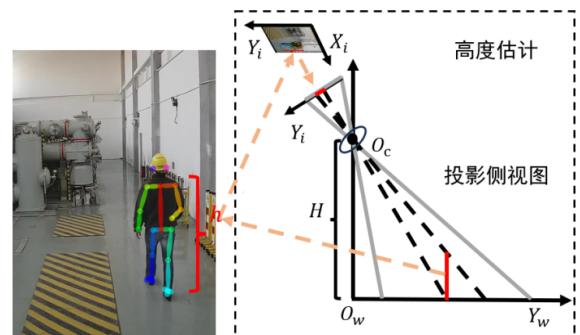


图 4 人员高度估计

1.3 空间转换

空间转换主要是基于前面提取得到的关键节点像素 (u, v) 和预测高度 h_{person} ，计算人员在世界场景空间的 $X_w O_w Y_w$ 平面的当下位置，可以通过公式 (8) 解得人员位置 (x, y) 。

$$\alpha \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = K [R|T] \begin{bmatrix} x \\ y_i \\ h_{\text{person}} \\ 1 \end{bmatrix}$$
(8)

其中， K 和 $[R|T]$ 分别是相机的内参和外参系数。

2 实验

在变电站作业场景开展真实实验,分为两个部分。第一部分针对高度估计进行实验测试,对于行走人员,取若干图像帧进行关键节点提取和分辨率映射计算人员高度,并与人员真实高度进行对比分析。第二部分利用本方法进行人员定位的测试,对于30 m内人员行走的基准直线和基于单帧图像的离散定位结果进行比较分析,高度估计实验采样共25张,结果如表1所示。

表1 人员高度估计结果

人员	真实身高/m	平均预测身高/m	平均误差/%	准确率/%
A	1.880 0	1.842 9	1.97	98.03
B	1.830 0	1.7877	2.31	97.69

由表1可见,人员高度估计误差大约在2%左右。相比于手工标记的图像分辨率映射方法,本方法的误差高出约1.5%,还未能够达到该方法的理论精度极限。认为这是由于位姿关键节点提取的误差决定的。在进行人员位姿估计时,关键节点的抖动看似不大,但是当进行计算时,首先必然会引起累计误差,其次是进行空间转换时关节抖动偏差在进行线性求解时引起函数解与真实位置偏移过大,都增大了定位误差。

而且值得注意的是,高度估计误差基本是低估。分析认为这是由于在通过人员的位姿进行人员高度估计时,二维图像上关键节点的提取时常并不能覆盖全部身高,因此造成人员高度估计偏低。

人员定位误差的散点图如图5所示,对于采样

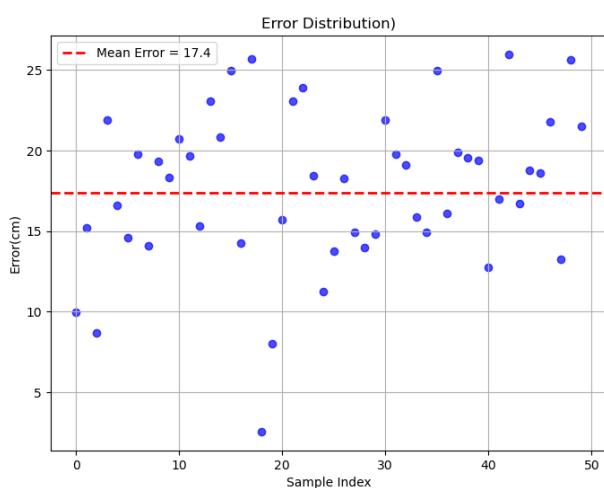


图5 人员定位误差分布

的50个样本,平均误差为17.4 cm,方差约为4.6 cm,最小误差为2 cm,最大误差为26.1 cm。发现人员定位结果似乎徘徊在17 cm上下,推测这可能是由于高度估计误差引起的基础误差,即在不提升高度估计精度的情况下很难进一步提升。当分析最小误差样本时,发现此时人员距离镜头处于较近位置,而此时的位姿识别头部高度更高,这样高度估计有了更高精度,从而造就了最小误差。

3 结束语

本文提出了一种基于像素分辨率映射的单目人员三维定位方法,通过提取人员关键节点二维信息、构建图像像素与世界空间的像素分辨率映射来实现人员高度估计。通过人员高度预测,结合空间转换实现在无特殊标定需求、无人员高度先验知识下的人员三维定位,其中图像分辨率映射方法也可用于单目场景下的目标三维重建。实验表明,本方法对于人员高度估计和人员定位,在摄像头已知的情况下存在可行性,高度估计由于关键节点覆盖不全普遍偏低,但高度估计精度约为98%,依然可观。人员定位方面相较于高度估计存在累计误差,平均误差约17.4 cm,最大误差26.1 cm。针对现存问题,计划优化更精确、稳定的人员关键节点提取模型,以提升该技术方案下的高度估计和人员定位精度。

参考文献:

- [1] 易能. 基于深度学习的双目视觉煤矿井下人员定位研究[D]. 中国矿业大学, 2023.
- [2] 王端, 刘世平, 王利军. 基于机器视觉的煤矿巷道人员定位研究[J]. 矿山机械, 2024, 52(1): 56-60.
- [3] 潘世一. 基于视觉定位技术的水力发电厂入侵人员UWB 定位方法[J]. 水力发电, 2023, 49(11): 109-114.
- [4] 张琨, 王旭航, 周立群, 等. 超宽带定位技术在火电厂输煤系统应用的试验研究[J]. 电力科技与环保, 2020, 36(5): 54-58.
- [5] 孙海渤, 岳红原, 王智睿, 等. 变电站三维可视化管控系统的设计与实现[J]. 工业控制计算机, 2020, 33(2): 97-99.
- [6] 缪坤坤. 基于 UKF-MTW 定位算法的变电站人员定位系统[D]. 安徽理工大学, 2023.
- [7] 刘哲, 董煊, 张德广, 等. 基于鲸鱼算法优化 DV-Hop 的人员定位算法[J]. 河北电力技术, 2022, 41(2): 85-89.
- [8] 薛灿, 王智. 基于声信道分析的电缆隧道人员定位技术[J]. 信息与控制, 2022, 51(2): 201-213.

- [9] 司福强. 5G+泛无线融合技术在地铁施工人员定位中的研究[J]. 铁路通信信号工程技术, 2024, 21(2): 89-95.
- [10] 范丽萍. 基于实时定位的建筑工人轨迹安全态势评价研究[D]. 中国矿业大学, 2023.
- [11] Zhuang Y, Yahng J. Smartphone-based indoor localization with bluetooth low energy beacons[J]. Sensors, 2016, 16(5): 596.
- [12] Xiong Z, Zhen Y S, Scalera A, et al. Enhancing WSN-based indoor positioning and tracking through RFID technology[C]/RFID technology (EURASIP RFID), 2012.
- [13] Lu H, Wang T, Ge F, et al. A robust UWB array localization scheme through passive anchor assistance[J]. China Communications (English Version), 2021, 18(4): 1-13.
- [14] Cardillo J, Sid-Ahmed M. 3-D position sensing using a passive monocular vision system[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 1991, 13(8): 809-813.
- [15] Zhang Z Y. Flexible camera calibration by viewing a plane from unknown orientations[C]/Proceedings of the Seventh IEEE International Conference on Computer Vision, Kerkyra, Greece, 1999: 666-673.
- [16] Zhang Y, Tangg S, Yu Z. Adaboost.RH: ensemble learning method for visual positioning[C]/Proceedings of the 16th National Conference on Signal and Intelligent Information Processing and Application, 2022.
- [17] Yu Z, Zhang Y, Feng J, et al. Vision-based personnel positioning framework in power operation environments [C]/2023 5th International Conference on Artificial Intelligence and Computer Applications (ICAICA), 2023: 127-131.
- [18] Tang S, Zhang Y, Feng J. Research on personnel positioning accuracy optimization technology for complex electric power environment[C]/2023 IEEE 5th International Conference on Civil Aviation Safety and Information Technology (ICCASIT), 2023: 338-342.
- [19] Mizuchi Y, Ogura T, Kim Y, et al. Accuracy evaluation of camera-based position and heading measurement system for vessel positioning at a very close distance [C]/Proceedings of the 15th International Conference on Control, Automation and Systems (ICCAS), Busan, Korea(South), 2015.
- [20] Han Y, Chung S, Yeh J, et al. Localization of RGB-D camera networks by skeleton-based viewpoint invariance transformation[C]/Proceedings of 2013 IEEE International Conference on Systems, Man, and Cybernetics, Manchester, UK, 2013.
- [21] Kim S, Lee J, Kim D, et al. Human localization based on the fusion of vision and sound system[C]/Proceedings of 2011 8th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI), Incheon, Korea(South), 2011.
- [22] Sun Y, Zhao K, Jian W, et al. Device-free human localization using panoramic camera and indoor map [C]/Proceedings of IEEE International Conference on Consumer Electronics-China, Guangzhou, China, 2016.
- [23] Jian Z, Zhu F, Tang L. Research on human body recognition and position measurement based on AdaBoost and RGB-D[C]/Proceedings of 2020 39th Chinese Control Conference (CCC), Shenyang, China, 2020.
- [24] Li X, Teng J, Zhai Q, et al. EV-human: human localization via visual estimation of body electronic interference[C]/Proceedings of IEEE Infocom, Turin, Italy, 2013.
- [25] Wu B X, Ay S U, Abdel-Rahim A. Trapezoid pixel array complementary metal oxide semiconductor image sensor with simplified mapping method for traffic monitoring applications[J]. Optical Engineering, 2018, 57(9).
- [26] Wu B X, Abdel-Rahim A, Ay S U. A trapezoid CMOS image sensor with 2% detection accuracy for traffic monitoring[C]/Proceedings of the 60th International Midwest Symposium on Circuits and Systems, Boston, USA, 2017: 1154-1158.
- [27] Wu B X, Ay S U, Abdel-Rahim A. Pedestrian height estimation and 3D reconstruction using pixel-resolution mapping method without special patterns[J]. Int J Autom Comput, 2019, 16: 449-461.
- [28] Dhall A, Chelani K, Radhakrishnan V, et al. LiDAR-camera calibration using 3D-3D point correspondences [J]. arXiv:1705.09785, 2017.
- [29] Gong X, Lin Y, Liu J. 3D LIDAR-camera extrinsic calibration using an arbitrary trihedron[J]. Sensors, 2013, 13(2): 1902.
- [30] Cao Z, Hidalgo G, Simon T, et al. OpenPose: realtime multi-person 2D pose estimation using part affinity fields[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 43(1): 172-186.

基于特征表达增强网络的无人机视角下车辆跟踪

刘琳洁¹, 后兴海^{2*}, 王辰¹, 刘畅¹, 盛智勇¹

1. 北方工业大学 信息学院, 北京 100144

2. 北方工业大学 电气与控制工程学院, 北京 100144

摘要: 随着遥感技术和人工智能的快速发展, 使用无人机遥感技术跟踪地面车辆是遥感领域的研究热点。然而, 无人机视角下的车辆跟踪仍存在诸多难题, 车辆尺度变化大、障碍物遮挡和相似车辆干扰均会影响跟踪的准确性和连续性。针对以上难题, 提出一种基于特征表达增强网络的无人机视角下车辆跟踪方法, 在 YOLOv5 中引入 EfficientFormerV2 网络, 并在卷积层构建条件参数卷积, 增强网络的特征表达能力; 构建双向 Hausdorff 距离匹配替换 DeepSORT 算法中的 IOU 距离匹配, 减少车辆 ID 切换次数。使用 DroneVehicle 数据集和 Berkeley DeepDrive Drone 数据集分别进行训练和验证, 实验结果表明, 该方法具有较好的跟踪性能, 平均精确度达到 91.8%, 跟踪准确率提高了 6.23%, ID 切换次数减少了 29%。

关键词: 无人机视频; 车辆跟踪; YOLOv5; DeepSORT; 条件参数卷积

文献标志码: A 中图分类号: TP391.43

Vehicle Tracking from UAV Perspective Based on Feature Representation Enhanced Network

LIU Linjie¹, HOU Xinghai², WANG Chen¹, LIU Chang¹, SHENG Zhiyong¹

1. College of Information, North China University of Technology, Beijing 100144, China

2. College of Electrical and Control Engineering, North China University of Technology, Beijing 100144, China

Abstract: With the rapid development of remote sensing technology and artificial intelligence, using UAV remote sensing technology to track ground vehicles is a research hotspot in the field of remote sensing. However, there are still many problems in vehicle tracking from the perspective of UAVs. Large changes in vehicle size, obstructions and similar vehicle interference will affect the accuracy and continuity of tracking. To solve the above problems, this paper proposes a method for vehicle tracking from the UAV perspective based on feature expression enhancement network. The YOLOv5 introduces EfficientFormerV2 network, and constructs conditional parameter convolution on the convolutional layer to enhance the feature expression capability of the network. Bidirectional Hausdorff distance matching is constructed to replace IOU distance matching in DeepSORT algorithm and reduce the number of vehicle ID switching. DroneVehicle dataset and Berkeley DeepDrive Drone dataset are respectively used for training and verification. Experimental results show that the proposed method has good tracking performance, with an average accuracy of 91.8%, tracking accuracy increases by 6.23%, and ID switching times reduce by 29%.

Key words: drone video; vehicle tracking; YOLOv5; DeepSORT; CondConv

基金项目: 北京市自然科学基金 (4232022); 北方工业大学毓秀创新项目 (2024NCUTYXCX119)。

作者简介: 刘琳洁 (2001—), 女, 硕士研究生, 研究方向为遥感图像智能处理; 后兴海 (1994—), 通信作者, 男, 博士研究生, 研究方向为智能信息处理; 王辰 (2002—), 女, 硕士研究生, 研究方向为遥感图像智能处理; 刘畅 (2001—), 女, 硕士研究生, 研究方向为遥感图像智能处理; 盛智勇 (1979—), 男, 高级实验师, 研究方向为电子信息、机器学习。

随着遥感技术和人工智能的快速发展,无人机在复杂场景下对地面车辆进行跟踪的能力已成为交通监管、公共安全与应急响应领域的关键技术支撑。无人机能够全方位地捕捉复杂场景下地面车辆的信息,并利用人工智能算法对海量数据进行处理,实现对地面车辆的精确定位与连续跟踪。

在无人机视角下车辆跟踪领域,研究者们提出了一系列创新性方法。Han 等人^[1]将快速判别相关滤波器与边界回归网络结合,消除基于无人机平台的车辆跟踪的累积漂移。Chen 等人^[2]利用宽残差网络从无人机航拍视频中提取车辆深度特征来实现精确跟踪并估计车辆速度。Maras 等人^[3]提出一种辅助粒子滤波跟踪算法,改进对车辆目标状态的估计,实现快速运动下的跟踪。但上述方法仍无法有效解决以下难题。首先,由于无人机飞行高度和视角的变化,同一辆车的尺度可能会发生显著变化,这种变化增加了车辆检测的难度。其次,车辆和障碍物之间可能会相互遮挡,降低车辆跟踪的准确性和连续性。此外,同一类别的车辆可能仅有车型和颜色之分,俯视成像视角下易造成相似车辆间相互干扰,分配给车辆的 ID 容易不断切换,增加了跟踪的难度。

针对上述问题,本文提出一种基于特征表达增强网络的无人机视角下车辆跟踪方法。首先,针对车辆尺度变化大和障碍物遮挡的问题,本文在 YOLOv5 中引入 EfficientFormerV2^[4]网络,并在卷积层构建条件参数卷积^[5]动态调整卷积核大小,增强网络的特征表达能力,提高车辆检测精度。其次,针对相似车辆的干扰导致车辆 ID 频繁切换的问题,构建双向 Hausdorff 距离匹配替代 DeepSORT 中的 IOU 匹配,通过考虑更多的特征信息来更好地区分无人机视角下的相似车辆,提高算法在相似车辆干扰下的跟踪精度和连续性。

1 方法

本文提出的方法在整体架构上进行创新性优化,在复杂的道路环境下能够准确检测出车辆的位置并连续地跟踪。在车辆检测阶段引入 EfficientFormerV2 网络,并在卷积层构建条件参数卷积,增强网络的特征表达能力;在车辆跟踪阶段构建双向 Hausdorff 距离匹配替代 IOU 匹配。总体流程如图 1 所示。

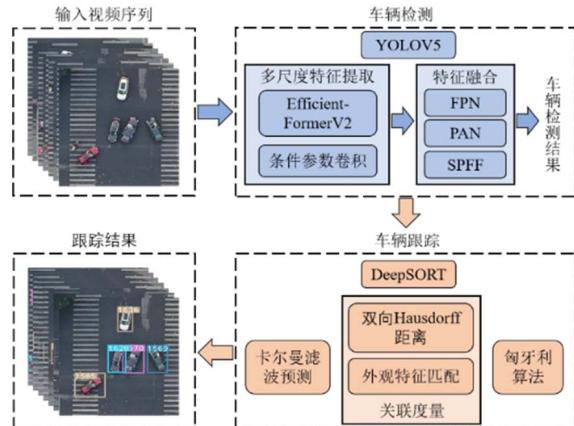


图 1 总体流程图

1.1 基于特征表达增强网络的车辆检测

YOLOv5^[6]采用 CSPDarknet53 网络作为特征提取网络,尽管在许多场景下表现优异,但在无人机视角下对地面车辆进行跟踪时仍存在一定局限性,在车辆尺度变化大和障碍物遮挡车辆的情况下检测效果不佳。因此,本文提出一种基于特征表达增强网络的车辆检测模型,引入 EfficientFormerV2 网络替换 YOLOv5 中的 CSPDarknet53 网络,同时在网络的卷积层构建条件参数卷积增强 EfficientFormerV2 网络的特征表达能力。EfficientFormerV2 网络结构如图 2 所示。

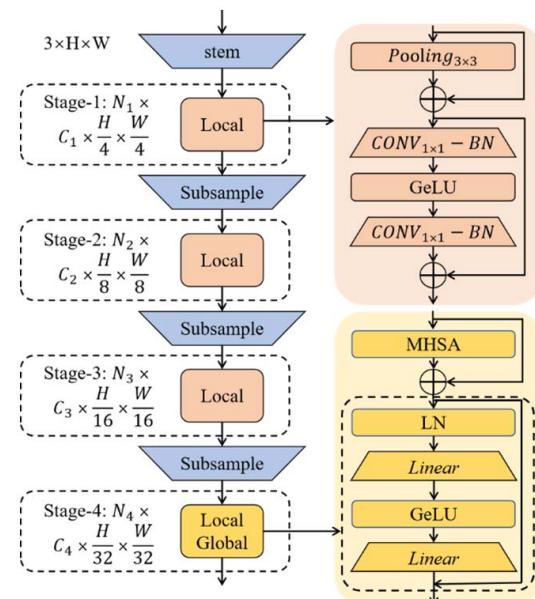


图 2 EfficientFormerV2 网络结构

EfficientFormerV2 采用四阶段分层设计的方法,这种方法在输入分辨率为 {1/4, 1/8, 1/16, 1/32}

处获得特征图，计算过程如下：

$$X_{i_{l=1}, j_{j=1}}^{B, C_j, \frac{H}{4}, \frac{W}{4}} = stem(X_0^{B, 3, H, W}) \quad (1)$$

式中， B 为批次大小， C 为通道尺寸， H 、 W 分别为特征的宽度和高度， X_0 为输入图像， X_{ij} 为第 i 层第 j 阶段的特征， $stem$ 表示卷积下采样操作。在四阶段分层设计中，前两个阶段捕获高分辨率下的局部信息，因此只采用相同的前馈网络（feed-forward neural network, FFN）处理每一层特征图。计算过程如下：

$$X_{i+1, j}^{B, C_j, \frac{H}{2^{j+1}}, \frac{W}{2^{j+1}}} = S_{i, j} \cdot FFN^{C_j, E_{i, j}}(X_{i, j}) + X_{i, j} \quad (2)$$

其中， $S_{i, j}$ 是一个可学习的层缩放因子，FFN 由两个属性构成：阶段宽度 C_j 和每个块的扩展比 $E_{i, j}$ 。在后两个阶段中，既使用了局部 FFN 也使用了全局多头自注意力块（multi-head self-attention, MHSA）。在上式的基础上，全局块的定义如下：

$$X_{i+1, j}^{B, C_j, \frac{H}{2^{j+1}}, \frac{W}{2^{j+1}}} = S_{i, j} \cdot MHSA(Proj(X_{i, j})) + X_{i, j} \quad (3)$$

为进一步提升模型的精度并减少模型复杂度，本文在 EfficientFormerV2 网络的卷积层构建条件参数卷积。条件参数卷积首先对输入的特征图进行分析，提取其不同的尺寸，并通过条件生成网络生成一组对应的条件权重，然后利用条件权重对基础卷积核进行加权组合，生成适应输入特征图属性的卷积核，并对输入特征图进行卷积操作，实现多尺度的特征提取与融合，提高特征表达能力。卷积过程如下式：

$$Output(x) = \sigma[(\alpha_1 W_1 + \alpha_2 W_2 + \dots + \alpha_n W_n) * x] \quad (4)$$

式中， α_i 为根据输入特征图 x 得到的条件权重， W_i 为预定义的基础卷积， σ 为非线性激活函数。

在前两个阶段利用局部 FFN 和条件参数卷积相结合，根据输入特征图的不同尺寸动态调整卷积核大小，以自适应提取精细的细节信息。在后两个阶段，结合局部 FFN、全局多头自注意力块以及条件参数卷积，进一步强化特征表达能力。条件参数卷积提供了更高的灵活性和自适应性，能够在不同尺度上自适应地捕捉丰富的特征信息，从而更好地应对无人机视角下车辆尺度变化大和障碍物遮挡

的问题。

1.2 基于双向 Hausdorff 距离的车辆跟踪

DeepSORT^[7] 算法采用 IOU 计算预测框和目标框的重合比例，以此作为评估预测结果与目标实际位置匹配程度的重要依据。然而，IOU 仅考虑了两个边界框的重叠区域，并未考虑目标的形状信息，不能有效处理相似车辆干扰导致的车辆 ID 频繁切换问题。因此，本文在 DeepSORT 算法中引入双向 Hausdorff 距离匹配，替换原有的 IOU 距离匹配，减少车辆 ID 切换的次数，提高车辆跟踪的精度和稳定性。

双向 Hausdorff 距离是一种用于衡量两个点集之间最大不匹配程度的度量方法，它定义为两个单向 Hausdorff 距离中的较大者，其中单向 Hausdorff 距离是从一个点集到另一个点集的所有点之间最小欧式距离中的最大值。具体而言，点集 A 和 B 的双向 Hausdorff 距离定义如下：

$$H(A, B) = \max[h(A, B), h(B, A)] \quad (5)$$

式中， $h(A, B)$ 表示点集 A 到点集 B 的单向 Hausdorff 距离， $h(B, A)$ 表示点集 B 到点集 A 的单向 Hausdorff 距离。 $h(A, B)$ 和 $h(B, A)$ 的计算公式如下，其中 $\| \cdot \|$ 代表两个点之间的欧氏距离。

$$\begin{aligned} h(A, B) &= \max_{a \in A} \min_{b \in B} \|a - b\| \\ h(B, A) &= \max_{b \in B} \min_{a \in A} \|b - a\| \end{aligned} \quad (6)$$

通过对每一对检测框和预测框计算它们的双向 Hausdorff 距离，并将得到的双向 Hausdorff 距离作为匹配的依据，通过设定阈值来判断检测框与预测框是否匹配。本文设定的最大双向 Hausdorff 阈值为 0.3，当计算得到的双向 Hausdorff 距离小于 0.3 时，则认为这一对检测框和预测框匹配成功；反之则匹配失败。根据匹配结果更新目标的轨迹信息，对匹配成功的检测框，使用卡尔曼滤波更新其轨迹状态；对于未匹配成功的检测框，创建新的轨迹。

双向 Hausdorff 距离综合考虑车辆之间的双向最大不匹配程度，提供更精细的匹配度量，增强模型对相似车辆的区分能力，提高跟踪的准确性和连贯性。

2 实验结果及分析

2.1 数据集构建

本文采用天津大学推出的大型无人机航拍车辆数据集 DroneVehicle^[8]中的 RGB 图像作为车辆检测的训练集，其中 RGB 图像共有 19 459 张，按照 7 : 2 : 1 的比例划分训练集、验证集和测试集；将车辆检测训练集中的所有检测目标截取出来，得到车辆跟踪的训练数据集。其次，采用 Berkeley DeepDrive Drone 数据集^[9]来验证车辆跟踪方法的性能，该数据集包含 20 个无人机航拍视频，记录了大量复杂非结构化场景。

2.2 实验平台和评价指标

本文的实验搭载了 4 块 NVIDIA GeForce GTX 2080Ti 独立显卡，配备 128 GB 内存和 8 TB 硬盘；在软件环境上，基于 Ubuntu18.04 LTS 操作系统，采用 Pytorch1.8.1 框架、Python3.7.13 和 CUDA10.1。实验策略上设置初始学习率为 0.01、batch-size 设置为 32，车辆检测和跟踪的训练总迭代次数均为 300。

为验证本文提出的方法在无人机视角下车辆跟踪任务的有效性，在车辆检测阶段，采用精确率、召回率、平均精确度作为检测算法性能的评估指标；在车辆跟踪阶段，采用多目标跟踪准确率 MOTA 和 ID 的切换次数 IDS_W 来评估改进后的车辆跟踪方法的性能。IDS_W 用于衡量跟踪过程中车辆 ID 发生错误更改的次数，MOTA 的计算公式如下：

$$MOTA = 1 - \frac{\sum_t (FN_t + FP_t + ID_{st})}{\sum_t GT_t} \quad (7)$$

2.3 实验结果分析

在车辆检测阶段，以 YOLOv5 为基准框架，通过消融实验证本文所提出的改进方法的有效性。消融实验结果如表 1 所示。

表 1 车辆检测消融实验结果 单位：%

Method	Precision	Recall	mAP
YOLOv5	87.6	81.5	84.7
Ours	90.4	87.2	91.8

相比较于原始的 YOLOv5 模型，基于条件参数卷积的特征表达增强网络 EfficientFormerV2 的

车辆检测模型精确率提高 2.8%，召回率提高 5.7%，平均精确度提高 6.7%，验证了改进算法的合理性和有效性。

在此基础上，验证基于双向 Hausdorff 距离匹配的 DeepSORT 算法的有效性，实验结果如表 2 所示。

表 2 车辆跟踪消融实验结果

Method	MOTA/%	IDS _W
YOLOv5-DeepSORT	87.55	1 305
Ours	93.78	926

由表 2 可知，基于改进后的 YOLOv5 模型，并在 DeepSORT 算法中引入双向 Hausdorff 距离匹配后，目标跟踪准确率提升 6.23%，ID 切换次数降低了 379 次。

为进一步展现本文所提出的方法在车辆尺度变化大、障碍物遮挡和相似车辆干扰情况下的有效性，本文选取同一视频中的第 1 867 帧、1 913 帧、1 952 帧、1 989 帧、2 061 帧和 2 135 帧的跟踪结果，如图 3 所示。

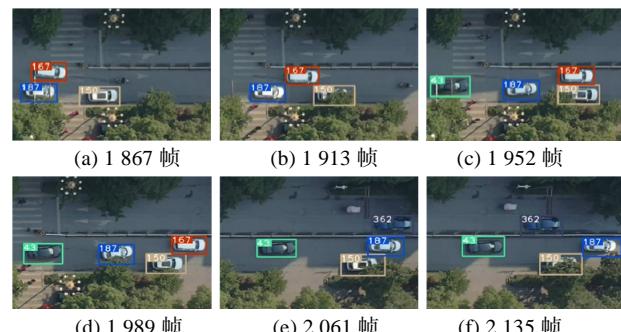


图 3 车辆跟踪结果图

从图 3 展现的跟踪结果可知，本文所提出的方法在无人机视角下车辆跟踪过程中展现出较好的准确性和连续性。视频中车辆由远及近驶来，发生了尺度变化，且存在颜色和形状相似的车辆，但依旧能够准确地锁定并持续跟踪车辆；此外，ID 为 150 的车辆在 1 867 帧中并未被遮挡，但在 1 913 帧、1 952 帧、1 989 帧和 2 061 帧中车辆信息均被部分遮挡，甚至在 2 135 帧中大部分车辆信息都被树木遮挡，但本文所提出的算法仍能够准确、连续地跟踪，证明了其在车辆尺度变化大、障碍物遮挡和相似车辆干扰下的有效性。

3 结论

本文提出一种基于特征表达增强网络的无人机视角下车辆跟踪方法。引入 EfficientFormerV2 网络替换 YOLOv5 中的 CSPDarknet53 网络，并在卷积层构建条件参数卷积，增强网络的特征表达能力，提高模型在车辆尺度变化大和障碍物遮挡下的车辆检测精度；其次，引入双向 Hausdorff 距离匹配替代 DeepSORT 算法中的 IOU 距离匹配，有效解决相似车辆干扰导致的车辆 ID 频繁切换的问题。实验结果表明，与原始算法相比，本文提出的方法在 Berkeley DeepDrive Drone 数据集上取得了更好的检测和跟踪结果，为无人机视角下的车辆跟踪提供一种理想的解决方案。

参考文献：

- [1] Han Y, Wang H, Zhang Z, et al. Boundary-aware vehicle tracking upon UAV[J]. Electronics Letters, 2020, 56(17): 873-876.
- [2] Chen Y, Zhao D, Er M J, et al. A novel vehicle tracking and speed estimation with varying UAV altitude and video resolution[J]. International Journal of Remote Sensing, 2021, 42(12): 4441-4466.
- [3] Maras B, Arica N, Ertuzun A. A robust vehicle tracking in low-altitude UAV videos[J]. Machine Vision and Applications, 2023, 34(5).
- [4] Li Y, Hu J, Wen Y, et al. Rethinking vision transformers for mobilenet size and speed[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023: 16889-16900.
- [5] Yang B, Bender G, Ngiam J, et al. CondConv: conditionally parameterized convolutions for efficient inference[J]. arXiv: 1904.04971, 2019.
- [6] Kumar S, Singh K S, Varshney S, et al. Fusion of deep sort and YOLOv5 for effective vehicle detection and tracking scheme in real-time traffic management sustainable system[J]. Sustainability, 2023, 15(24).
- [7] Wojke N, Bewley A, Paulus D. Simple online and real-time tracking with a deep association metric[C]//I EEE International Conference on Image Processing, 2017.
- [8] Sun Y, Cao B, Hu Z Q. Drone-based RGB-infrared cross-modality vehicle detection via uncertainty-aware learning[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2022, 32(10): 6700-6713.
- [9] Wu F, Wang D, Hwang M, et al. Decentralized vehicle coordination: the Berkeley DeepDrive drone dataset[J]. arXiv: 2209.08763, 2022.

(上接第 162 页)

- [10] HUA Y. Estimating two-dimensional frequencies by matrix enhancement and matrix pencil[J]. IEEE Transactions on Signal Processing, 1992, 40(9): 2267-2280.
- [11] HUA Y. A pencil-MUSIC algorithm for finding two-dimensional angles and polarizations using crossed dipoles[J]. IEEE Transactions on Antennas and Propagation, 1993, 41(3): 370-376.
- [12] HAARDT M, ZOLTOWSKI M D, MATHEWS C P, et al. 2D unitary ESPRIT for efficient 2D parameter estimation[C]//1995 International Conference on Acoustics, Speech, and Signal Processing. IEEE, 1995: 2096-2099.
- [13] YANG Z, LI J, STOICA P, et al. Sparse methods for direction-of-arrival estimation[M]//Academic Press Library in Signal Processing, Volume 7. Academic Press, 2018: 509-581.
- [14] YAN H, CHEN T, WANG P, et al. A direction-of-arrival estimation algorithm based on compressed sensing and density-based spatial clustering and its application in signal processing of MEMS vector hydrophone[J]. Sensors, 2021, 21(6): 2191.
- [15] WANG Q, ZHAO Z, CHEN Z. Fast compressive sensing DOA estimation via ADMM solver[C]//2017 IEEE International Conference on Information and Automation (ICIA). IEEE, 2017: 53-57.
- [16] WIPF D P, RAO B D. Sparse Bayesian learning for basis selection[J]. IEEE Transactions on Signal processing, 2004, 52(8): 2153-2164.
- [17] XU S, KOOIJ B J, YAROVOY A. Joint Doppler and DOA estimation using (ultra-) wideband FMCW signals[J]. Signal Processing, 2020, 168: 107259.
- [18] ZHU D, LI Y, ZHU Z. A keystone transform without interpolation for SAR ground moving-target imaging[J]. IEEE Geoscience and Remote Sensing Letters, 2007, 4(1): 18-22.
- [19] WAX M, SHAN T J, KAILATH T. Spatio-temporal spectral analysis by eigenstructure methods[J]. IEEE Transactions on Acoustics, Speech, and Signal Processing, 1984, 32(4): 817-827.
- [20] WANG H, KAVEH M. Coherent signal-subspace processing for the detection and estimation of angles of arrival of multiple wide-band sources[J]. IEEE Transactions on Acoustics, Speech, and Signal Processing, 1985, 33(4): 823-831.
- [21] NANNURU S, GEMBA K L, GERSTOFFT P, et al. Sparse Bayesian learning with multiple dictionaries[J]. Signal Processing, 2019, 159: 159-170.