

DOI: 10.19650/j.cnki.cjsi.J2615268

融合姿态引导与频域补偿的中远距离 双视角视线测量方法*

杨傲雷^{1,2}, 梅尹颀¹, 杨帮华¹, 苗中华¹

(1. 上海大学机电工程与自动化学院 上海 200444; 2. 上海市电站自动化技术重点实验室 上海 200444)

摘要:针对中远距离(1.0~3.0 m)视线测量中因成像尺度退化与大角度自遮挡导致的精度衰退问题,提出一种融合姿态引导与频域尺度补偿的双视角视线测量方法。首先,为突破单一近场数据集的验证局限,构建了中远距离双视角时空对齐数据集,完整保留了真实自然状态下眼-头动态耦合的复杂机理,为视线测量算法提供可靠的验证基准。其次,设计了头部姿态引导的动态特征调节模块,利用连续的 6D 姿态表征构建高维空间注意力语义掩码,通过自顶向下的先验约束,隔离并抑制因侧视透视畸变与物理遮挡引入的空间噪声。随后,为克服中远距离下眼部特征尺度急剧退化与拓扑细节丢失的瓶颈,提出尺度自适应高频补偿模块,通过双眼像素间距动态感知物理观测距离,并以此驱动尺度嵌入,结合频域二维离散余弦变换,自底向上重构在光学低通衰减中受损的高频眼动边界特征。最后,引入基于不确定性路由的多视角协同融合策略,在保障多视角空间几何物理一致性的基础上,该策略能够对受噪声干扰严重的单侧降质数据进行动态门控降权,实现跨异构视角特征的深度互补。实验结果表明,该方法在自建受控验证集与 ETH-XGaze 通用数据集上的平均角度误差分别降至 3.52° 与 2.75°,展现出更优的测量精度与鲁棒性。

关键词:双视角视线测量;姿态引导;频域补偿;多模态特征融合;中远距离

中图分类号: TP391 TH86 **文献标识码:** A **国家标准学科分类代码:** 510.4050

Mid-to-long-range dual-view gaze measurement via pose guidance and frequency-domain compensation

Yang Aolei^{1,2}, Mei Yinjie¹, Yang Banghua¹, Miao Zhonghua¹

(1. School of Mechatronic Engineering and Automation, Shanghai University, Shanghai 200444, China;

2. Shanghai Key Laboratory of Power Station Automation Technology, Shanghai 200444, China)

Abstract: To address accuracy degradation caused by imaging scale decay and large-angle self-occlusion in mid-to-long-range (1.0~3.0 m) gaze measurement, a dual-view method integrating pose guidance and frequency-domain compensation is proposed. First, to overcome the limitations of near-field datasets, a mid-to-long-range dual-view spatiotemporally aligned dataset that preserves natural eye-head dynamic coupling mechanisms is constructed to provide a reliable benchmark. Second, a head pose-guided feature selection module is designed. By leveraging continuous 6D pose representations to generate spatial attention masks, it uses top-down prior constraints to isolate and suppress spatial noise from perspective distortion and occlusion. Subsequently, to tackle severe scale degradation and topological detail loss, a scale-adaptive high-frequency reconstruction compensation module is proposed. It dynamically infers the physical observation distance via inter-ocular pixel spacing to drive scale embeddings, and employs the 2D discrete cosine transform to bottom-up reconstruct attenuated high-frequency ocular boundary features in a bottom-up manner. Finally, an uncertainty-aware weight learning fusion strategy is introduced. While ensuring spatial geometric consistency, it dynamically penalizes unilaterally degraded views to achieve deep cross-view feature complementation. Experimental results demonstrate that the method reduces the mean angular error to 3.52° and 2.75° on the custom and ETH-XGaze datasets, respectively, exhibiting superior measurement accuracy and robustness.

Keywords: dual-view gaze measurement; pose guidance; frequency-domain compensation; multi-modal feature fusion; mid-to-long-distance

0 引 言

三维视线测量作为解析人类视觉感知与行为意图的核心手段,在智能座舱监控、远程医疗辅助及人机交互等领域展现出重要的工程应用价值^[1]。随着物理交互空间由传统的桌面近距离向 1.0~3.0 m 的中远距离范围拓展,非接触式视线测量系统^[2]已成为当前仪器仪表与交互测量领域的研究热点。相比于限制运动自由度的穿戴式设备或应用场景受限的近场固定式眼动仪^[3],中远距离非接触测量能够在不干扰受试者自然行为状态的前提下,提供连续的空间注视数据流。

现有的视线测量系统主要分为穿戴式与单视角固定式两类。穿戴式眼动仪虽具备较高的测量精度,但受试者需长期佩戴头戴装置^[4],极易引发视觉疲劳并限制其运动自由度,难以满足长时程、无感知的测控任务部署需求。单视角固定式系统虽实现了非接触测量,但在中远距离的自由交互空间下却面临严峻的物理测量瓶颈:1) 观测距离的增加导致眼部图像分辨率呈现多尺度退化现象,高频边缘细节严重丢失;2) 非受限交互中不可避免的头部姿态变化,极易引发面部及眼部区域严重的透视畸变与自遮挡^[5]。

为克服单视角测量在中远距离场景下存在的深度歧义、局部遮挡及对姿态变化敏感等难题,将三维几何建模、多视角协同感知及跨模态信息融合引入视线解算框架,已成为提升系统在复杂非受限环境中测量鲁棒性与工程适应能力的重要途径。相较于传统单目方案,多视角测量体制能够跨空间观测角度提供丰富的冗余几何约束,弥补因单侧自遮挡造成的信息缺失,进而提升三维空间视线测量的精确度与一致性。

在融合姿态先验引导方面,针对后期特征融合极易引发流形空间冲突的物理局限,基于特征解耦与多尺度注意力的视线估计方法 DMA-Gaze^[6]提出了一种基于连续正交旋转表征的姿态解耦框架,从测地线距离层面抑制了姿态参数空间中的奇异性与不连续问题。基于地标引导与交叉注意力的视线估计网络 LGNet^[7]利用条件变分自编码器构建空间几何的生成式先验,通过双流交叉注意力与门控重加权机制,实现了视线关键点与表观特征的高维空间对齐,从而主动抑制了因大角度引发的局部遮挡噪声。

在多视角协同测量领域,引入空间观测冗余是缓解物理遮挡与深度歧义的有效手段。无监督域适应的双视角视线估计方法 UVAGaze^[8]通过构建双相机阵列初步改善了单视角下的视觉盲区问题,双视角视线估计模型 DV-Gaze^[9]基于双视角特征融合策略,进一步提升了极端姿态下的解算稳定性。为突破传统测量系统静态标定与固定特征权重的局限,多视角注视目标估计网络(multi-

view gaze-target estimation, MVGT)^[10]引入了不确定性感知机制,通过联合建模各空间观测分支与数据不确定性,动态量化不同视角的信噪比与可信度水平,并在高维融合阶段自适应调节各空间视角的贡献权重。

在高维特征提取与多模态补偿方面,单视角视线解算极易受环境光照波动及中远距离成像带来的有效像素衰减影响。为此,基于多头注意力融合与交互的视线估计网络 MAFI-Gaze^[11]利用多头交叉注意力机制处理侧视状态下双眼表观的非对称交互,通过面眼交互模块引导网络聚焦高信息熵的眼周区域。自适应特征融合网络 AFF-Net^[12]从特征归一化机制入手,设计自适应组归一化模块,依据全局上下文动态调节局部眼部特征分布,有效抑制了因受试者个体差异、光照突变及局部尺度波动引起的特征漂移。基于双向交叉注意力的视线估计模型 BiCross-Gaze^[13]构建了双向交叉注意力通路,在宏观面部空间拓扑与微观眼部纹理细节之间建立深度交互,实现了多尺度特征的协同表征。相比于单向增强策略,该结构更充分地挖掘了局部观测与全局信息间的内在物理关联。此外,双分支特征融合视线估计网络 Gaze-Former^[14]利用无注意力机制(attention free transformer, AFT)与门控单元构建平行融合架构,在降低系统计算开销的同时提升了无约束交互场景下的鲁棒性。通过整合面部空间拓扑、微观眼部纹理及三维姿态等异构信息,现有多模态融合技术构建了具备物理冗余度的高维解算体系,从底层信号处理层面有效补偿了中远距离视线测量的感知盲区与尺度退化问题^[15]。

综上所述,针对中远距离交互空间中面临的成像尺度退化与自遮挡等物理测量瓶颈,本文提出了一种融合头部姿态引导与频域补偿的双视角视线测量方法。该方法首先构建了姿态引导的动态特征调节机制,利用连续空间姿态先验生成注意力掩码,主动抑制大角度偏航带来的透视畸变与遮挡噪声;其次,为应对距离骤增引起的视觉细节丢失,设计了尺度自适应高频补偿模块,通过动态解算双眼物理间距并结合频域二维离散余弦变换与条件卷积,实现高频眼动边界特征的自适应重构;最后,引入基于不确定性感知路由的多视角融合网络,实现了对单侧降质视角的动态降权与跨空间特征的协同互补。此外,本文构建了面向中远距离非受限交互的双视角时空对齐数据集,为本测量系统的模型寻优与性能验证提供了三维视线物理基准。

1 问题描述及方法架构

1.1 视线测量问题定义与坐标系拓朴

在数学上,真实视线方向被定义为一个单位矢量: $\mathbf{g} \in \mathbf{S}^2 \subset \mathbf{R}^3$,代表一条从设定的空间参考原点 O (如头

部几何中心或双眼中心)出发,指向注视目标的物理射线。本文提出的双视角视线测量方法,旨在学习一组映射函数,将因远距离而退化的图像输入,投影到该视线流形上。系统输出的测量值规范化定义为三维视线方向矢量 $\mathbf{G} \in \mathbf{S}^2$,该输出矢量与真实视线 $\mathbf{g}$ 严格同构于单位球面流形,确保系统解算结果在三维空间中具有刚性的方向约束与物理测量意义。

本文构建的中远距离双视角视线测量系统的物理架构与坐标系拓扑关系如图1所示。

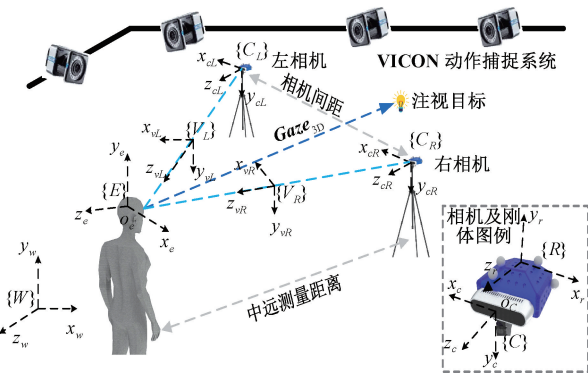


图1 中远距离双视角视线测量系统空间几何模型

Fig.1 Spatial geometric model of mid-to-long-distance dual-view gaze measurement system

该测量系统主要由 VICON 光学动作捕捉系统与异构布置的双视角测量节点(左、右相机)协同构成。为实现异构视觉传感器与物理真实空间的对齐,系统内定义了以下核心物理与几何坐标系:1)全局参考坐标系 $\{W\}$,建立在真实物理空间中的世界坐标系 (x_w, y_w, z_w) ,将系统中的 VICON 坐标系定义为世界坐标系,作为所有空间节点与视线的绝对标定基准;2)目标特征坐标系 $\{E\}$,以受试者为中心的头部/眼部参考坐标系 (x_e, y_e, z_e) ,其原点 O_e 为三维视线矢量 $\mathbf{Gaze}_{3D}$ 的物理起点,蓝色射线即表征了由受试者眼部中心指向空间注视目标的真实视线方向;3)物理观测坐标系 $\{C_L\}$ 与 $\{C_R\}$,分别代表左相机与右相机的独立光心坐标系,图中标注的“相机间距”与“中远测量距离”直观反映了本系统大基线、长摄距的工程测量特性。4)视线虚拟归一化坐标系 $\{V_L\}$ 与 $\{V_R\}$,为剥离物理相机的外部安装参数并消除非对称透视畸变,系统依托几何解算构建的理想正交观测空间。以左支路为例,其虚拟 Z 轴 z_{vL} 被严格约束为从物理光心指向眼部原点 O_e 的空间矢量。

此外,图1右下角局部放大图例描述了单侧视觉测量节点的位姿追踪与标定机制。通过在相机后部固联光学反光标识点,构建了独立追踪的刚体坐标系 $\{R\}$ 。在系统初始化阶段,通过求解物理相机坐标系 $\{C\}$ 与刚体坐标系 $\{R\}$ 之间的恒定空间转换关系,可将双视角相机

的局部观测数据高频、实时地映射至 VICON 主导的全局世界坐标系 $\{W\}$ 中,从而为视线解算模型提供具备物理约束的三维真值基准。

1.2 视线测量的不确定性与几何共面约束

在中远距离非受限交互下,由于成像尺度锐减、极限姿态诱发的自遮挡以及环境光照的非均匀衰减,测量系统面临严重的数据偶然不确定性。这种物理层面的信号退化是不可逆的,无法单纯依靠增加网络深度来消除。因此,本系统摒弃传统的确定性点估计范式,建立一种基于偶然不确定性感知的概率评估机制,以量化并隔离劣质视角的观测噪声,从而在源头上保障融合基准的可靠性。

此外,中远距离三维视线测量本质上是一个高度敏感的空间射影几何问题。随着受试者与传感器阵列之间物理距离的增加,系统面临严峻的空间几何误差放大效应。微小的像素级特征偏移或角度解算偏差,在长基线投射下会导致三维空间中的视线射线发生严重的几何退化与发散。为抑制这种由距离放大的拓扑发散,需要引入先验的物理结构强约束。

本文采用双路异构视觉传感器阵列,从三维射影几何原理出发,引入多视角共面约束(coplanarity constraint)先验。由于异构观测相机在同一时刻凝视的均为同一物理注视目标,其各自解算出的三维视线射线在理想运动学空间中,必须严格交于同一点并保持共面状态。本方法将利用这一刚性的宏观空间拓扑结构作为自监督信号,在模型特征寻优阶段收缩高维参数解空间,抑制违背物理透视规律的离群预测,确保多源异构信号的测量一致性。

1.3 测量系统总体架构

本文提出的视线测量系统总体架构如图2所示。该系统采用双支路异构输入设计,旨在通过多模态特征补偿与不确定性路由策略,克服中远距离姿态下的测量衰退问题。整个处理过程可分为3个阶段:

1)异构视觉感知与特征提取(左侧):系统首先获取由空间中不同观测视角(相机1与相机2)捕获的受试者图像,并在物理空间完成刚体变换($\mathbf{T}$)与时空对齐。随后,左、右视图图像独立并行送入“多模态特征提取层”,系统利用深层卷积网络同步抽取出蕴含全局面部拓扑与局部眼部细节的高维表现特征。

2)多模态特征解码与动态调制(中部):在特征解耦与深层映射阶段,提取到的初级特征被分为两条独立的补偿路径。(1)利用分离出的眼部微观特征,输入“尺度自适应高频超分补偿”模块(scale-adaptive high-frequency super-resolution, SASR),针对中远距离成像模糊问题,进行高频边缘细节的超分辨率重构;(2)将解算出的头部

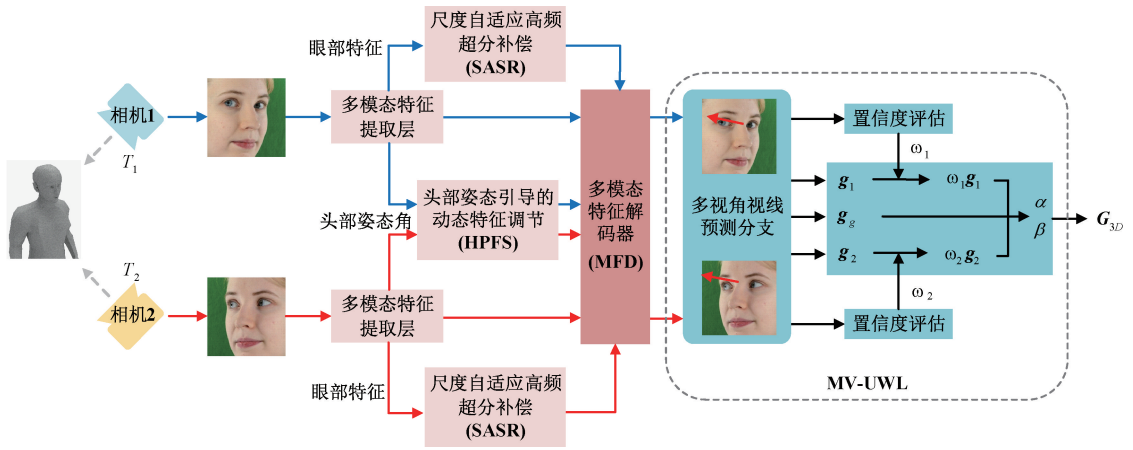


图2 多模态自适应融合的中远距离视线测量架构

Fig. 2 Multimodal adaptive fusion model for mid-to-long-range gaze measurement

姿态角输入“头部姿态引导的动态特征调节”模块 (head pose-guided feature selection, HPFS), 生成空间特征掩码以主动抑制遮挡噪声。经过上述先验调制与超分补偿后的高质量特征, 被同步送入多模态特征解码器 (multimodal feature decoder, MFD) 进行深层的特征重组与语义对齐。

3) 基于不确定性路由的跨视角融合 (右侧): 经过 MFD 解码后的高维特征最终进入多视角不确定性权重学习模块 (multi-view uncertain weight learning, MV-UWL)。多视角视线预测分支分别回归出左侧局部视线 $Gaze_1$ 、右侧局部视线 $Gaze_2$ 以及全局协同视线 $Gaze_g$ 。同时, 独立并行的置信度评估模块对单侧降质视角的不确定性进行量化, 输出动态权重 ω_1 与 ω_2 。最终, 系统通过软门控机制 α, β 将置信度加权后的单侧预测值与全局预测值进行跨空间互补求和, 输出具备物理一致性与鲁棒性的三维视线矢量 $Gaze_{3D}$ 。

2 视线测量模型与算法机理

2.1 异构视角下的空间几何解耦与归一化

为消除多摄像机物理位置差异带来的异构观测, 以及受试者自由移动引起的透视畸变, 系统需在特征感知前, 将采集的双视图图像与视线真值同步映射至规范化的虚拟相机坐标系。设真实相机坐标系的原点为 O_c , 眼部中心的三维空间坐标为 $\mathbf{e} \in \mathbf{R}^3$ 。

首先, 定义虚拟相机的主光轴方向 $\mathbf{z}_v$ 为由光心指向眼部中心的单位矢量。为消除平面外的滚转干扰, 引入绝对物理参考“上方向”矢量 $\mathbf{u} = [0, 1, 0]^T$ 。通过施加正交化约束依次构建虚拟相机的 $\mathbf{x}_v$ 轴与 $\mathbf{y}_v$ 轴, 从而推导出正交旋转矩阵 $\mathbf{R} = [\mathbf{x}_v, \mathbf{y}_v, \mathbf{z}_v]^T$, 其中, $\mathbf{z}_v = \mathbf{e} / \|\mathbf{e}\|_2$, $\mathbf{x}_v = (\mathbf{u} \times \mathbf{z}_v) / \|\mathbf{u} \times \mathbf{z}_v\|_2$, $\mathbf{y}_v = \mathbf{z}_v \times \mathbf{x}_v$ 。

随后, 依据透视投影的几何相似原理, 结合设定的近场标准参考摄距 d_s 与实际物理观测距离 $\|\mathbf{e}\|_2$, 构建尺度缩放矩阵 $\mathbf{S}$ 。针对中远距离面部及眼部像素区域小、特征易丢失的物理局限, 该矩阵在维持二维平面宽高比不变的前提下, 对远距离的微弱特征进行规范化的空间放大与重采样, 如式 (1) 所示。

$$\mathbf{S} = \text{diag}\left(1, 1, \frac{d_s}{\|\mathbf{e}\|_2}\right) \quad (1)$$

综合上述几何操作, 作用于二维图像平面的透视变换归一化矩阵 $\mathbf{M}$ 表示为:

$$\mathbf{M} = \mathbf{K}_v \mathbf{S} \mathbf{R} \mathbf{K}_c^{-1} \quad (2)$$

式中: $\mathbf{K}_c$ 与 $\mathbf{K}_v$ 分别为真实相机与设定的虚拟标准相机内参矩阵。该映射不仅裁切出空间对齐的眼部感兴趣区域, 亦实现了三维视线标签的同步坐标系转换。这一过程可有效剥离相机外部参数的几何耦合, 降低网络在中远距离下回归弱特征分布的拟合负担。

2.2 融合姿态先验与频域补偿的双维特征调制机制

中远距离视线测量面临两大核心瓶颈: 1) 头部大角度偏转引发的透视畸变与物理遮挡; 2) 观测距离增加导致的眼部高频特征多尺度模糊。传统的视线测量网络大多依赖于深度卷积的隐式特征提取, 面对这种耦合退化时往往陷入特征纠缠。针对这一底层缺陷, 本文提出了一种自顶向下几何约束与自底向上信号重构相协同的双维特征干预机制。该机制旨在信号提取的早期, 主动引入三维物理朝向先验与离散频谱分析, 从数学机理上解耦并补偿上述退化过程。

首先, 原始局部面部图像张量 $\mathbf{F} \in \mathbf{R}^{H \times W \times 3}$ 经残差网络压缩, 映射为包含全局上下文与微观细节的深层语义特征矢量 $\mathbf{f}_l = \Phi_{\text{ResNet}}(\mathbf{F}) \in \mathbf{R}^{512 \times 1}$ 。

在空间拓扑维度 (自顶向下的几何约束), 为阻断遮挡畸变带来的结构性噪声, 本文引入姿态引导调制执行

流形层面的动态门控。该模块不依赖隐式的视觉纹理,而是将确定的头部三维朝向物理量 $\mathbf{P}_0 \in \mathbf{R}^3$ 经级联的流形投影网络转化为高维空间注意力掩码 $\mathbf{h}_s = f_4 \circ f_3 \circ f_2 \circ f_1(\mathbf{P}_0)$, 其中,第 i 层映射为 $f_i(\mathbf{x}) = \text{ReLU}(\mathbf{W}_i \mathbf{x} + \mathbf{b}_i)$ 。随后,利用 Sigmoid 函数将 $\mathbf{h}_s$ 归一化为门控系数,对初始特征 $\mathbf{f}_i$ 施加通道级权重加权,输出经姿态校准的优化表征 $\mathbf{f}'_i$, 如式(3)所示。

$$\mathbf{f}'_i = \mathbf{f}_i \odot \sigma(\mathbf{h}_s) \quad (3)$$

式中: $\odot$ 为 Hadamard 乘积。此过程在数学上相当于引入了一个物理约束的信息瓶颈,主动抑制畸变区域的特征响应,从宏观上保障复杂姿态下的特征拓扑连续性。

在信号尺度与频域维度(自底向上的微观补偿),针对上述空间门控可能带来的微观高频细节衰减,本文并行构建了跨尺度频域边缘重构机制。该机制并非简单的图像锐化,而是建立在正交频谱空间中的信号靶向修复。设标准近距基准面的标定眼距像素值为 D_{ref} ,当前帧中双眼外眼角的欧氏像素距离 D_{eo} ,解算出表征分辨率退化的尺度因子 s ,如图3所示。随后采用多层感知机将其非线性映射为尺度感知嵌入矢量 $\mathbf{e}_s \in \mathbf{R}^{64 \times 1}$,如式(4)所示。

$$\mathbf{e}_s = \text{ReLU}(\mathbf{W}_{s2} \cdot \text{ReLU}(\mathbf{W}_{s1} \cdot s + \mathbf{b}_{s1}) + \mathbf{b}_{s2}) \quad (4)$$

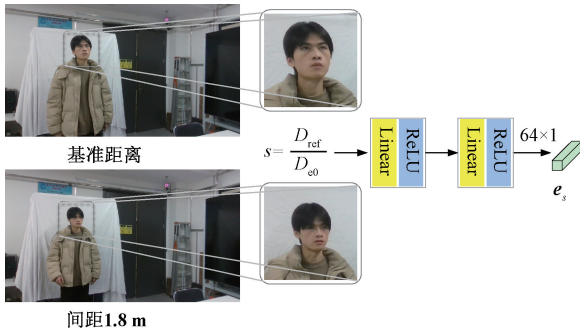


图3 基于面部生物学拓扑的尺度退化测算示意图

Fig. 3 Schematic diagram of scale degradation estimation based on facial topology

在信号修复阶段,对眼部空间域信号 $\mathbf{F}_{\text{eye}}$ 执行二维离散余弦变换(two-dimensional discrete cosine transform, 2D-DCT)^[16],将其无损映射至正交频谱域,如式(5)所示。

$$\mathbf{F}_{\text{freq}} = 2\text{D-DCT}(\mathbf{F}_{\text{eye}}) \quad (5)$$

引入理想高通滤波掩码 $\mathbf{M}_{\text{hp}}$ 滤除低频光照与肤色干扰后,通过二维离散余弦逆变换(two-dimensional inverse discrete cosine transform, 2D-IDCT)重构出高频残差边缘特征 $\mathbf{F}_{\text{hr}}$,如式(6)所示。

$$\mathbf{F}_{\text{hr}} = 2\text{D-IDCT}(\mathbf{F}_{\text{freq}} \odot \mathbf{M}_{\text{hp}}) \quad (6)$$

最后,以尺度感知矢量 $\mathbf{e}_s$ 驱动动态路由策略重组卷积参数,完成对 $\mathbf{F}_{\text{hr}}$ 的跨尺度补偿。上述双维机制在物理

机理上实现了协同,可以提升模型在中远距离与极端姿态下的特征不变性。

2.3 融合显式几何先验的多模态特征解码器

本文提出多模态特征解码器,旨在改善纯视觉表征的感知局限。该机制将受试者眼部的二维归一化坐标作为显式几何锚点,通过跨模态交叉注意力算法,建立视觉纹理流形与空间几何拓扑之间的深层语义对齐。依托这种刚性的空间先验,系统能够指导视觉特征进行动态权重分配,主动过滤结构性噪声。为了消除图像像素流形与数值坐标空间之间的语义鸿沟,系统主要依托图4所示的多模态特征提取层对多源输入数据进行处理。

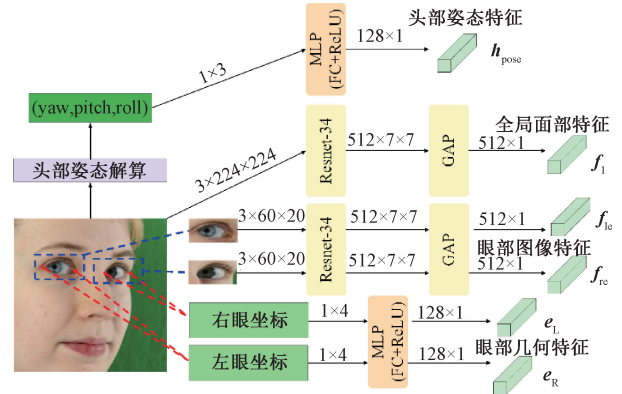


图4 多模态特征提取层结构与数据流转示意图

Fig. 4 Architecture and data flow of the multi-modal feature extraction layer

系统采用样本与计算重分配人脸检测器(sample and computation redistribution for efficient face detection, SCRFD)检测器^[17]锁定受试者的面部边界。在视觉感知支路,对多源输入进行非线性空间映射与维度统一,并将降维后的全局面部特征 $\mathbf{f}_1$ 、局部左/右眼特征($\mathbf{f}_{\text{le}}$ 与 $\mathbf{f}_{\text{re}}$)以及三维头部姿态矢量 $\mathbf{h}_{\text{pose}}$ 进行通道级拼接,并通过非线性映射层生成融合姿态先验的单侧联合视觉特征 $\mathbf{f}_{\text{IL}}$ 与 $\mathbf{f}_{\text{IR}}$:

$$\mathbf{f}_{\text{IL}} = \sigma(\mathbf{W}_{\text{fuse}}[\mathbf{f}_1; \mathbf{f}_{\text{le}}; \mathbf{h}_{\text{pose}}] + \mathbf{b}_{\text{fuse}}) \quad (7)$$

$$\mathbf{f}_{\text{IR}} = \sigma(\mathbf{W}_{\text{fuse}}[\mathbf{f}_1; \mathbf{f}_{\text{re}}; \mathbf{h}_{\text{pose}}] + \mathbf{b}_{\text{fuse}}) \quad (8)$$

其中, $[\cdot; \cdot]$ 表示特征矢量的级联拼接; $\sigma(\cdot)$ 为非线性激活函数。考虑到人类左右眼部生理结构的高度对称性,此处采用孪生权重共享机制。即左眼支路与右眼支路严格共享同一组 $\mathbf{W}_{\text{fuse}}$ 与 $\mathbf{b}_{\text{fuse}}$, 以确保跨模态特征融合过程的物理一致性,同时有效抑制模型对单侧眼部特征的过拟合。随后,将双眼特征在特征维度上堆叠为 $\mathbf{f}_{\text{ILR}}$,并经线性变换投影至紧凑的跨模态子空间,生成视觉查询特征矩阵 $\mathbf{Q} = \mathbf{W}_v \mathbf{f}_{\text{ILR}} + \mathbf{b}_v$ 。

在空间几何支路,并行通道对二维归一化眼部坐标执行深层拓扑编码。设左、右眼的二维绝对中心坐标分

别为 $\mathbf{p}_L \in \mathbf{R}^2$ 与 $\mathbf{p}_R \in \mathbf{R}^2$ 。该坐标组经由多层感知机进行非线性升维,转化为高维几何特征矢量 $\mathbf{e}_L$ 与 $\mathbf{e}_R$:

$$\mathbf{e}_L = \text{ReLU}(\mathbf{W}_{L2} \cdot \text{ReLU}(\mathbf{W}_{L1}\mathbf{p}_L + \mathbf{b}_{L1}) + \mathbf{b}_{L2}) \quad (9)$$

$$\mathbf{e}_R = \text{ReLU}(\mathbf{W}_{R2} \cdot \text{ReLU}(\mathbf{W}_{R1}\mathbf{p}_R + \mathbf{b}_{R1}) + \mathbf{b}_{R2}) \quad (10)$$

将两者拼接构建表征双眼整体空间结构的联合位置特征矩阵 $\mathbf{E}_{LR} = [\mathbf{e}_L, \mathbf{e}_R]$ 。在后续的交叉注意力机制中,该纯几何特征矩阵同步充当键矩阵 $\mathbf{K}$ 与值矩阵 $\mathbf{V}$,即 $\mathbf{K} = \mathbf{V} = \mathbf{E}_{LR}$ 。

最终,网络利用眼部坐标的几何拓扑为视觉表征施加空间分布约束。通过计算视觉查询矩阵 $\mathbf{Q}$ 与空间位置键矩阵 $\mathbf{K}$ 的特征相关性,对包含拓扑信息的几何值矩阵 $\mathbf{V}$ 执行动态加权聚合:

$$\mathbf{H}_{\text{attn}} = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_h}}\right) \mathbf{V} \in \mathbf{R}^{128 \times 2} \quad (11)$$

式中: d_h 为调节内积方差的缩放维度系数。该交叉注意力运算的核心物理实质在于:利用图4中几何支路输出的无偏特征 $\mathbf{e}_L, \mathbf{e}_R$ 作为可靠的空间锚点,对受畸变或遮挡污染的视觉隐式表征执行主动修正与特征补偿。最终输出的深度协同融合特征 $\mathbf{H}_{\text{attn}}$,有效剥离几何冗余与视觉噪声,为末端的多视角视线回归模块提供增强信噪比的输入流形。

2.4 基于不确定性感知路由的多视角融合策略

本文设计了多视角不确定性权重学习模块,通过量化各视角的不确定性,实现跨异构空间特征的动态路由与互补融合,其数据流转架构如图5所示。

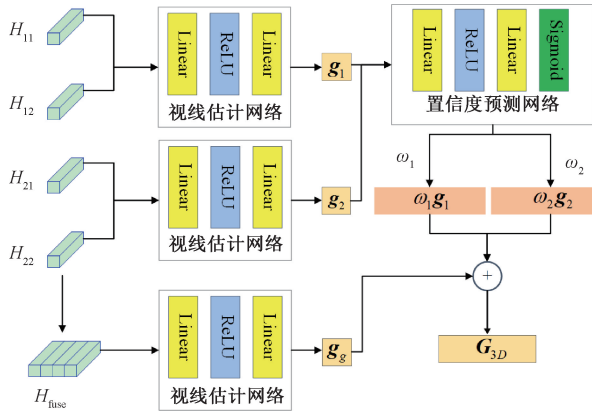


图5 多视角权重学习模型的数据流转架构

Fig. 5 Data flow architecture of the multi-view weight learning model

在信号分配阶段,系统首先将经过跨模态协同的高维融合特征分解为左视图特征组 $\mathbf{H}_{11}, \mathbf{H}_{12}$ 与右视图特征组 $\mathbf{H}_{21}, \mathbf{H}_{22}$ 。其中,特征张量 $\mathbf{H}_{ij}(i, j \in \{1, 2\})$ 分别指代物理视图与左右侧眼部视图,是特定视图下的单眼局部细节与该视图全局先验(含面部纹理、眼部绝对坐标及三维头部姿态)融合后的综合表征。基于此特征拓扑,系统

构建了3个并行的视线评估模型,两条分别处理单侧视觉流的独立评估分支,以及一条汇聚双侧特征的全局协同分支。

在局部单视角分支中,各独立评估模型在回归出初步的局部视线预测矢量 $\mathbf{g}_i \in \mathbf{R}^3, i \in \{1, 2\}$ 的同时,将其隐式特征馈入由全连接层与 ReLU (rectified linear unit) 激活函数交替构成的置信度预测网络,并在末端利用 Sigmoid 函数将特征映射为连续的置信度标量 $\omega_i \in [0, 1]$:

$$\omega_i = \sigma(\mathbf{W}_\omega \mathbf{v}_i + \mathbf{b}_\omega), i \in \{1, 2\} \quad (12)$$

式中: $\mathbf{v}_i$ 为置信度网络提取的局部特征状态矢量; $\mathbf{W}_\omega$ 与 $\mathbf{b}_\omega$ 为可学习的参数。标量 ω_i 客观量化当前单侧视角受光照衰减或遮挡影响的物理不确定程度,充当后续特征融合的动态路由权重。

在全局协同分支中,系统将所有视角的局部特征张量沿通道维度进行拼接,构建出包含全局上下文的融合特征块 $\mathbf{X}_{\text{fuse}}$,并据此独立回归出具备多视角结构先验的全局协同视线矢量 $\mathbf{g}_g \in \mathbf{R}^3$ 。

在最终的融合解算阶段,系统通过软性门控机制,将全局双视角先验预测 $\mathbf{g}_g$ 与经置信度加权后的各局部独立预测 $\mathbf{g}_i$ 进行动态平滑组合。为确保输出矢量在三维几何空间中的刚性方向属性,最终的三维视线输出矢量 $\mathbf{G}_{3D} \in \mathbf{S}^2$ 需经过 L_2 范数归一化处理:

$$\tilde{\mathbf{G}}_{3D} = \alpha \mathbf{g}_g + \beta \sum_{i=1}^2 \omega_i \mathbf{g}_i, \mathbf{G}_{3D} = \frac{\tilde{\mathbf{G}}_{3D}}{\|\tilde{\mathbf{G}}_{3D}\|_2} \quad (13)$$

其中, α 和 β 为控制全局先验与局部微观预测相对重要性的自适应学习参数。为保障最终融合的三维视线测量矢量在欧氏空间中的物理合法性,该参数须服从凸组合数学约束,即满足 $\alpha, \beta \in [0, 1]$ 且 $\alpha + \beta = 1$ 。该不确定性感知路由机制可确保在单侧视觉发生物理衰减时,系统能够自动阻断高噪支路的信号传播(即 $\omega_i \rightarrow 0$),将解算置信度平滑过渡至全局分支与优势视角,从而从底层架构上保障人体视线测量的鲁棒性。

2.5 基于物理拓扑约束的多目标混合损失函数

系统的损失函数 L_{total} 由多视角角度损失 L_{ang} 与三维共面几何损失 L_{geo} 共同构成。针对批次尺寸为 N 的训练样本,其整体联合优化目标定义为:

$$L_{\text{total}} = \frac{1}{N} \sum_{n=1}^N (L_{\text{ang}} + L_{\text{geo}}) \quad (14)$$

在角度回归层面,设样本的真实视线方向为归一化矢量 $\mathbf{g} \in \mathbf{S}^2$,系统设计了自适应多视角角度一致性评估函数 L_{ang} ,以迫使各支路预测逼近物理真值:

$$L_{\text{ang}} = \lambda_1 \|\mathbf{g} - \mathbf{g}_g\|_2^2 + \lambda_2 \sum_{i=1}^2 \left(\frac{\|\mathbf{g} - \mathbf{g}_i\|_2^2}{2\omega_i^2} + \frac{1}{2} \ln \omega_i^2 \right) \quad (15)$$

式中: $\mathbf{g}_g$ 全局协同分支预测的视线矢量; $\mathbf{g}_1$ 与 $\mathbf{g}_2$ 分别为

左、右单视角分支的独立预测矢量。 λ_1 和 λ_2 为平衡权重系数。 ω_i 为由 MV-UWL 模块预测的单侧置信度标量。当某侧视角发生遮挡时,网络倾向于预测较大的不确定性 ω_i ,从而动态衰减该降质样本在总体损失中的惩罚权重,并在公式后半段通过对数正则项防止权重无限发散。

在空间几何约束层面,为有效收缩高维参数求解空间并剔除异面离群值,系统引入基于三维矢量混合积的共面一致性正则化项 L_{geo} 。针对同一注视目标,由系统左、右局部支路与全局支路解算的三维视线射线必须表现出高度的线性相关性并严格共面。该约束通过最小化 3 个矢量张成的平行六面体体积来执行:

$$L_{\text{geo}} = \eta |(\mathbf{g}_1 \times \mathbf{g}_2) \cdot \mathbf{g}_g|^2 \quad (16)$$

其中, η 为几何惩罚缩放系数。通过在系统参数寻优阶段引入该物理拓扑强约束,模型能有效过滤违背三维射影几何规律的局部测量噪声,提升测量系统在复杂中远距离场景下的鲁棒性与物理可解释性。

3 实验平台与数据集构建

3.1 中远距离双视角测量系统与数据时空对齐机制

数据集采集硬件由 VICON 光学运动捕捉系统与双视角视觉感知系统组成。其中,动捕系统通过实时追踪目标上的红外反光 Marker 阵列,为系统提供毫米级的三维空间位置与刚体姿态基准。视觉感知系统采用两台 RealSense 相机,相机背面与特制支架进行刚体连接。两台相机对称布置于受试者前方左右两侧的特定基线上,形成双侧协同观测视角,如图 6 所示。



图 6 数据集采集场景

Fig. 6 Dataset acquisition scenario

多源异构数据的深度融合高度依赖于严格的时空标定与对齐机制。在空间几何对齐方面,系统采用手眼标定方法,求解两台 RealSense 相机坐标系 $\{C\}$ 与 VICON 全局世界坐标系 $\{W\}$ 之间的外参矩阵。设旋转矩阵为 $\mathbf{R}_W^C \in \mathbf{R}^{3 \times 3}$, 平移矢量为 $\mathbf{T}_W^C \in \mathbf{R}^{3 \times 1}$, 则世界坐标系下的任意三维物理点 $\mathbf{P}_W$ 映射至相机坐标系的刚体变换模型可表示为 $\mathbf{P}_C = \mathbf{R}_W^C \mathbf{P}_W + \mathbf{T}_W^C$ 。

在时间序列同步方面,系统采用网络时间协议(network time protocol, NTP)强制同步视觉传感器与 VICON 主控机。在数据动态采集阶段,系统以 RealSense 相机底层的硬件触发信号为时间基准,提取时间戳严格匹配的动捕帧序列,从而实现跨模态数据流的毫秒级时序协同。

原始异构数据的预处理与真值标定包含面部拓扑提取、姿态解算与真值矢量生成 3 个环节。首先,针对视觉图像,系统锁定受试者的面部边界并回归出 106 个面部语义关键点,基于提取的内外眼角与轮廓坐标,定位双眼感兴趣区域并执行标准化空间裁剪。在头部姿态解算阶段,为克服欧拉角在大角度偏转时固有的万向节死锁缺陷,本文引入 6DRepNet360 架构^[18]回归连续的六维旋转矩阵,经正交化处理后解算出受试者三维头部姿态参数。

三维视线物理真值的标定基于严格的空間两点定线原理。系统结合相机内参及二维关键点反投影矩阵,定位受试者的三维眼球几何中心作为视线原点;同时,以光学动捕系统追踪的 3D 注视目标(LED 靶点)实际物理坐标作为视线终点。在计算由原点指向终点的单位方向矢量后,系统利用包含尺度缩放与旋转对齐的透视变换矩阵,将该原始物理矢量统一映射至视线虚拟归一化坐标系 $\{V_L\}$ 与 $\{V_R\}$ 。

3.2 实验数据集及分布特征

为评估所提模型的测量边界与真实泛化性能,本文采用“受控物理几何基准”与“通用环境泛化基准”相结合的数据集联合策略:

1) 自建中远距离双视角受控物理数据集:在三维视线测量中,透视畸变引发的空间几何方差对系统精度的破坏远超局部纹理差异。由于当前开源数据集中缺乏覆盖连续中远距离的双视角数据,本文针对性地构建了该数据集。为孤立验证空间几何退化效应,采集时呈现实验室环境中的单一受控场景。其物理测距分布覆盖 1.0~3.0 m 的动态交互空间,姿态分布包含大范围自由头部偏转。虽然该集合仅包含 4 名受试者,但通过高频连续采样获取了 25 000 组样本,构建了稠密、连续的三维几何数据集。部分采集样例如图 7 所示。

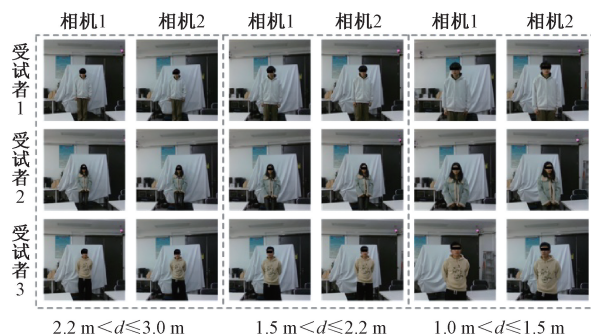


图 7 自建数据集例图

Fig. 7 Sample images from the self-built dataset

2) ETH-XGaze 通用泛化评估数据集:为弥补受控物理验证集在受试者规模与场景多样性上的局限,本文引入大规模 ETH-XGaze 数据集开展跨域联合验证。实验抽取了其中 10 名非重叠受试者(即测试集与训练集在物理身份上严格正交隔离)的约 30 000 张测试图像。这些数据涵盖了多样化的受试者面部生理拓扑、大角度姿态以及复杂的光照衰减分布。依托该数据集,可用于评估本文双维特征干预机制在复杂场景下的协同测量可靠性与环境适应能力。

3) 主流多视角三维视线数据集属性对比:表 1 列举了本数据集与当前主流多视角/3D 视线数据集在多维物理属性上的差异。

表 1 多视角数据集对比表

Table 1 Comparison of multi-perspective datasets

名称	规模	视线范围 (俯仰/偏航)	动态 头部	采集距 离/m
ColumbiaGaze ^[19]	5 880	±15°/±10°	否	~2
UT-multiview ^[20]	64 000	±50°/±36°	否	~0.6
ETH-XGaze ^[21]	1 083 492	±120°/±70°	否	~0.6
本文数据集	25 000	±120°/±80°	是	1~3

当前视线估计领域的基准数据集存在两大局限。(1)测距分布失衡:以 UT-multiview 与 ETH-XGaze 为代表的大规模数据集,其物理摄距高度集中于 0.5~1.0 m 的近场范围,缺乏 2.0 m 以上中远距离尺度退化的样本。尽管 ColumbiaGaze 涉及了 2.0 m 测距,但其样本规模小且仅包含 5 个离散头部朝向。(2)眼-头耦合机理的缺失:主流多视角数据多采用固定受试者头部、利用半球形相机阵列被动采集的策略。这种静态约束范式无法反映自然交互状态下眼球转动与头部自由运动之间复杂的几何耦合。

本文自建数据集将高频细节退化与大姿态遮挡双重退化效应显性化。它迫使测量算法的评估标准从“近场理想纹理拟合精度”向“恶劣边界下的抗崩溃能力”转移。同时,依托工业级 VICON 系统提供真值,数据集实现了毫秒级双视觉流信号与三维物理空间姿态的严格对齐。这一时空对齐属性,为研究和验证跨视角几何共面约束、联合注意力校准等抗退化测量算法,提供了真实可靠的物理标定基准。

4 测量模型评估与对比实验分析

4.1 评估指标与实验参数设置

在三维人体视线测量任务中,评估解算模型精度的核心标准是系统输出的预测视线矢量与真实物理视线矢

量之间的角度偏差。为确保实验评估的客观性及与国际同类研究的横向可比性,本系统采用平均角度误差(mean angular error, MAE)作为量化测量性能的评价指标。给定测试集中的样本索引 i ,设其经由动捕系统标定的三维视线物理真值为 $\mathbf{g}_i$,网络模型输出的三维视线测量矢量为 $\hat{\mathbf{g}}_i$,对于总容量为 n 的测试样本集,MAE 定义为:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n \arccos \left(\frac{\hat{\mathbf{g}}_i \cdot \mathbf{g}_i}{\|\hat{\mathbf{g}}_i\|_2 \|\mathbf{g}_i\|_2} \right) \quad (17)$$

式中: $\|\cdot\|_2$ 表示矢量的 L_2 范数。MAE 数值越小,表明测量系统的回归精度越高、特征映射能力越强。

尽管当前跨数据集横向评估时常采用 MAE,但其仅能反映系统误差的全局中心趋势,数学平均过程中,容易平滑并掩盖由远距离光学分辨率下降、高频特征退化及极端姿态遮挡所引发的偶发性失真(即长尾误差)。为了缓解这一指标带来的评估盲区,本文在后续的实验分析中,并未局限于全局 MAE 的单点对比,而是引入了针对极端头部姿态的细粒度区间评估策略。通过量化极端退化边界下的局部角度误差,力求在保持基准可比性的同时,客观揭示测量系统在多尺度物理退化下的鲁棒性。

本实验代码基于 PyTorch 框架构建,模型训练、参数寻优与实时推理解算均在一台搭载 NVIDIA GeForce RTX 4070 GPU 的计算平台上完成。在测量系统的参数标定与特征寻优阶段,实验采用自适应矩估计(adaptive moment estimation, Adam)优化器驱动网络的梯度下降与权重更新。初始基础学习率设定为 1×10^{-4} ,并引入余弦退火衰减策略以动态调节学习步长,保障模型在寻优后期的平稳收敛。网络训练的批次大小为 64,总 Epochs 设定为 60 轮。在定义的多目标混合损失函数中,角度回归损失权重 λ_1, λ_2 与几何共面惩罚权重 η 均采用网格搜索策略确定。此外,所有视觉传感器的原始输入图像在输入主干网络之前,均经过严格的空分辨率对齐与像素级标准归一化处理,以消除异构传感器底层物理响应的量纲差异。

4.2 测量系统性能验证与鲁棒性分析

基线模型的选择遵循“任务定义一致性”与“实验可复现性”原则。尽管近年来涌现了诸如 GMGaze、GA3CE 等工作,但此类工作多聚焦于单视角估计或注视目标预测等,且部分关键算法尚未开源。由于输入模态、多视角标定协议及解算目标空间的异构性,直接进行定量横向对比容易引入系统性评估偏差。所以,本文从受试者泛化能力、尺度退化抗性以及极端姿态容错率 3 个维度展开量化分析,并选取 UVAGaze 与 DVGaze 作为对比基线。

1) 受试者独立交叉验证与跨域实验分析

在自建双视角数据集上,本文采用受试者独立交叉

验证 (leave-one-person-out, LOPO) 策略,以评估测量系统对未知个体生理特征的泛化能力。表 2 对比了各模型在 LOPO 策略及公开数据集上的 MAE。

表 2 各模型在 LOPO 验证策略下的 MAE
Table 2 Results for the MAE of each model under LOPO

模型	平均角度误差/(°)	
	本文数据集	ETH-XGaze
UVAGaze	4.96	3.32
DVGaze	5.32	3.65
本文方法	3.52	2.75

测试结果表明,本文方法在自建集上的 MAE 低至 3.52°,相较于同属双视角架构的 UVAGaze (4.96°) 与 DVGaze (5.32°),其测量误差分别降低了 29.0% 与 33.8%。此外,在 ETH-XGaze 测试集上,本方法同样展现出 2.75°的精度。该结果从系统层面印证了多模态协同感知机制在抑制个体表观差异、提升标定稳定性方面的优势。

2) 物理测距延伸下的尺度退化抗性分析

为验证尺度自适应高频超分补偿模块对高频信号丢失的修复能力,实验依据受试者与相机阵列的绝对物理距离 d ,将测试集划分为 1.0~1.5 m、1.5~2.2 m、2.2~3.0 m 3 个区间。实验结果如表 3 所示。

表 3 各模型在不同采集距离下的 MAE
Table 3 Results for the MAE of each model under different acquisition distances

模型	平均角度误差/(°)		
	1.0 m≤ d ≤ 1.5 m	1.5 m< d ≤ 2.2 m	2.2 m< d ≤ 3.0 m
UVAGaze	4.35	5.02	5.82
DVGaze	4.58	5.45	6.35
本文方法	3.35	3.56	3.75

随着测距向 3.0 m 延伸,传统基准模型因深层特征发生严重的尺度退化,其测量误差显著发散 (UVAGaze 退化至 5.82°);相比之下,本文方法依托频域内的高频细节重构机制,在 2.2 m 以外的远距衰减区维持了 3.75°的解算水平,验证了系统具备一定的尺度退化抗性。

3) 偏转姿态下的空间几何鲁棒性分析

针对大角度头部偏转引发的物理自遮挡与传感器非对称信号盲区问题,实验按照受试者的真实偏航角 (y_{aw}) 范围进行了量化分级评估。如表 4 所示,在 $|y_{aw}|>60^\circ$ 的侧视工况下,单侧相机捕获的眼部图像发生严重的拓扑

畸变或完全遮挡。

表 4 各模型在不同头部姿态下的 MAE 对比
Table 4 Results for the MAE of each model under different head poses

模型	平均角度误差/(°)		
	$ y_{aw} \leq 30^\circ$	$30^\circ < y_{aw} \leq 60^\circ$	$ y_{aw} > 60^\circ$
UVAGaze	4.28	5.05	5.86
DVGaze	4.52	5.48	6.45
本文方法	2.98	3.55	4.62

本文方法由于在前端引入了融合姿态先验的门控机制以物理隔离畸变特征流,并在后端级联了不确定性感知路由以实现异构视角间解算置信度的主动转移,最终使得模型在极端姿态区间的测量误差被约束在 4.62°。相较于采用静态特征拼接的传统融合策略,本文的架构在面临非对称物理遮挡时展现出了一定的容错能力。

4.3 核心模块消融实验与机理分析

本节构建了 3 种模型变体,并在自建双视角数据集与 ETH-XGaze 公开数据集上执行了消融评估。各变体的配置及解算误差如表 5 所示。

表 5 核心功能模块的消融实验结果
Table 5 Results of the ablation experiments for core modules

模型变体	本文自建数据集/(°)	ETH-XGaze/(°)
Model w/o HPFS	4.02	3.08
Model w/o SASR	4.75	2.86
Model w/o MV-UWL	3.95	2.92
本文方法	3.52	2.75

1) 姿态门控机制的噪声阻断作用:移除融合姿态先验的门控机制 (w/o HPFS) 后,本文数据集上的误差增加了 0.50°。这表明在中远距离非受限交互中,头部大幅偏转引发的结构性噪声很严重。HPFS 的缺失导致系统丧失了在前端物理隔离畸变特征流的能力,致使劣质表观特征直接影响了后端的流形空间。

2) 尺度补偿机制的远距重构作用:当剥离尺度自适应高频补偿模块 (w/o SASR) 时,模型在本文中远距离数据集上的精度损失攀升了 1.23°;但该变体在常规近距离的 ETH-XGaze 数据集上误差上升了 0.11°。这一跨域表现差异,不仅反向印证了本文自建数据集在中远距尺度退化上的严苛性,更从机理上证实了 SASR 模块在网络中补偿高频视觉信号、抵抗远距离透视退化的作用。

3) 动态路由机制的容错作用:去除不确定性感知路由模块 (w/o MV-UWL),即退化为传统视角的平均静态

融合后,系统精度出现了 0.43° 的下降。这表明当异构传感器网络面临非对称遮挡与单侧信号盲区时,MV-UWL能够依据不确定性实现解算权重的动态转移,能够保障多传感器协同测量体系具备一定的容错特性。

此外,对于驱动网络几何一致性的静态损失超参数 η ,实验表明系统对其具体取值并不具有脆弱的敏感性。在网络参数的寻优阶段,当 $\eta \in [0.1, 0.3]$ 时,系统在各评估集上的量化误差波动较小,展现出宽泛的平稳有效区间。仅当 η 的量级过大时,才会因过度刚性约束而削弱对高频视觉残差的学习。本系统最终取经验值 $\eta=0.2$ 以平衡几何约束与表征学习,确保中远距离测量的稳定性。

4.4 真实动态场景的在线实测与定性评估

为进一步验证在真实物理环境下的工程实测效能,本文构建了在线实时三维视线测量原型系统。实验在受试者自由交互的典型工况下展开定性评估验证,其测试现场如图 8 所示。



图 8 真实场景下的双视角测量系统配置图

Fig. 8 Configuration diagram of the dual-view measurement system in real dynamic scenarios

测量系统在不同空间测距尺度与头部偏转姿态下的三维视线效果如图 9 所示。

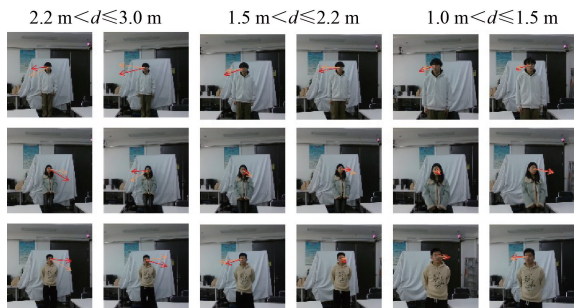


图 9 不同空间尺度与姿态下的三维视线测量结果示例

Fig. 9 Results of 3D gaze measurement under different spatial scales and head poses

其中,两组方向箭头之间的夹角直观地反映了三维视线真实值与预测值之间的角度偏差。

观测结果表明:在 $1.0 \sim 1.5$ m 的近中距交互区间内,由于系统能够获取眼部区域表现图像,模型预测的橙色射线与红色物理真值射线呈现出较高的空间重合度,证实了系统在常规尺度下具备较好的空间解算精度。而当物理测量基线大幅延伸至 2.2 m 以外的远距时,尽管受限于成像物理极限,眼部有效离散像素锐减并伴随高频纹理的透视模糊,测量矢量的绝对角误差出现了预期范围内的增加。但其整体三维空间指向依然与注视靶点的物理真值保持了一定的一致性,系统模型并未因严重的尺度退化或局部自遮挡而引发视线的发散与方向性崩溃。

综上所述,本文系统通过深度耦合姿态门控过滤机制、尺度自适应高频补偿机制以及基于不确定性感知路由的跨视角动态融合策略,不仅直观佐证了前文量化消融数据的真实可靠性,更验证了该测量架构在复杂工程应用现场的可行性。

5 结 论

本文面向中远距离非接触式视线测控需求,提出一种基于多模态协同与频域补偿的双视角三维视线测量方法。针对成像尺度退化与非对称遮挡痛点,系统构建基于姿态先验的门控机制主动隔离透视畸变与遮挡噪声;引入尺度自适应高频补偿,利用频域变换抑制高频视觉信号衰减;并依托不确定性感知路由实现跨视角动态容错与协同互补。相关实验表明,该方法能有效抑制人机交互工况下的视线拓扑发散,展现出一定的空间解算精度与鲁棒性。未来工作将探索引入连续帧时序追踪机制和融合红外主动成像等多模态信息,以进一步提升测量系统在极端光学退化环境下的全局泛化能力。

参考文献

[1] Wang J Z, Wang T F, Xu B J, et al. Accurate eye tracking from dense 3D surface recon-structions using single-shot deflectometry [J]. Nature Communications, 2025, 16(1) : 1-12.

[2] 程鸣, 严运兵. 驾驶员视线区域自适应半监督学习标定方法[J]. 电子测量与仪器学报, 2023, 37(12) : 136-142.

Cheng Ming, Yan Yunbing. Adaptive semi-supervised learning calibration method for driver's gaze region[J]. Journal of Electronic Measurement and Instrument, 2023, 37(12) : 136-142.

[3] Zhu H T, Yang H, Xu S Q, et al. Frequency-encoded eye tracking smart contact lens for human-machine interaction [J]. Nature Communications, 2024, 15: 3588.

- [4] 潘新星, 汪辉, 陈灵, 等. 基于卷积目标检测的3D眼球追踪系统深度估计[J]. 仪器仪表学报, 2018, 39(10): 241-248.
Pan Xinxing, Wang Hui, Chen Lin, et al. Convolution object detection based depth estimation of 3D eye-tracking system[J]. Chinese Journal of Scientific Instrument, 2018, 39(10): 241-248.
- [5] 胡佳辉, 陆永华, 张进海, 等. 基于机器学习的头部自由视线追踪方法及其在电动病床端的应用[J]. 仪器仪表学报, 2021(12): 101-109.
Hu Jiahui, Lu Yonghua, Zhang Jinhai, et al. Head-free line-of-sight tracking method based on machine learning and its application in electric beds[J]. Chinese Journal of Scientific Instrument, 2021(12): 101-109.
- [6] Chen H H, Liu H J, Lan S Y, et al. DMAgaze: Gaze estimation based on feature disentanglement and multi-scale attention[PP/OL]. V2. (2025-05-25)[2026-05-10]. <https://arxiv.org/abs/2504.11160><https://doi.org/10.48550/arXiv.2504.11160>.
- [7] Xu G H, Zhang X Y, Zhao W L, et al. Landmark-guided gaze estimation via conditional keypoint generation and cross-attention fusion[J]. Information, 2026, 17(3): 224.
- [8] Liu R C, Lu F. UVAGaze: Unsupervised 1-to-2 views adaptation for gaze estimation[C]//Proceedings of the AAAI Conference on Artificial Intelligence, 2024, 38(4): 3693-3701.
- [9] Cheng Y H, Lu F. DVGaze: Dual-view gaze estimation[C]//2023 IEEE/CVF International Conference on Computer Vision, 2023: 20578-20584.
- [10] Miao Q M, Golani V R, Xu J Y, et al. Multi-view gaze target estimation[C]//2025 IEEE/CVF International Conference on Computer Vision, 2025: 5371-5381.
- [11] Li C L, Li F F, Zhang K, et al. Gaze estimation network based on multi-head attention, fusion, and interaction[J]. Sensors, 2025, 25(6): 1893.
- [12] Bao Y W, Cheng Y H, Liu Y F, et al. Adaptive feature fusion network for gaze tracking in mobile tablets[C]//2020 25th International Conference on Pattern Recognition. IEEE, 2021: 9936-9943.
- [13] Wang Y F, Li C L. Gaze estimation model based on bidirectional cross-attention[C]//2025 7th International Conference on Frontier Technologies of Information and Computer. IEEE, 2025: 808-817.
- [14] Zhang D L, Li C L. Gaze estimation network with two-branch feature fusion[C]//Second International Conference on Image Processing, Intelligent Control, and Computer Engineering, 2025, 13942: 76-84.
- [15] Bai Z, Gao C, Shou M Z. EVOLVE-VLA: Test-time training from environment feedback for vision-language-action models[PP/OL]. V1. (2025-12-16)[2026-05-10]. <https://doi.org/10.48550/arXiv.2512.14666>.
- [16] Zeng Q Y, Hui B, Liu Z J, et al. A method combining discrete cosine transform with attention for multi-temporal remote sensing image matching[J]. Sensors, 2025, 25(5): 1345.
- [17] Vu H N, Nguyen M H, Pham C. Masked face recognition with convolutional neural networks and local binary patterns[J]. Applied Intelligence, 2022, 52: 5497-5512.
- [18] Hempel T, Abdelrahman A A, Al-Hamadi A. Toward Robust and unconstrained full range of rotation head pose estimation[J]. IEEE Transactions on Image Processing, 2024, 33: 2377-2387.
- [19] Smith B A, Yin Q, Feiner S K, et al. Gaze Locking: Passive eye contact detection for human-object interaction[C]//Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology, 2013: 271-280.
- [20] Sugano Y, Matsushita Y, Sato Y, et al. Appearance-based gaze estimation with online calibration from mouse operations[J]. IEEE Transactions on Human-Machine Systems, 2015. 45(6): 750-760.
- [21] Zhang X C, Park S, Beeler T, et al. ETH-XGaze: A large scale dataset for gaze estimation under extreme head pose and gaze variation[C]//European Conference on Computer Vision, 2020: 365-381.

作者简介



杨傲雷(通信作者), 2004年于湖北工业大学获得学士学位, 2009年于上海大学获得硕士学位, 2012年于英国女王大学获得博士学位, 现为上海大学副教授, 主要研究方向为机器人与视觉控制、计算机视觉与感知定位等。

E-mail: aolei@shu.edu.cn

Yang AoLei (Corresponding author) received his B. Sc. degree from Hubei University of Technology in 2004, his M. Sc. degree from Shanghai University in 2009, and his Ph. D. degree from Queen's University Belfast, UK, in 2012. He is currently an associate professor at Shanghai University. His main research interests include robotics and vision control, computer vision, and perception localization.