

# 文献综述：基于 RGB 颜色空间的红外偏振人脸识别

## 1. 文献信息

- **作者：** 汪方斌，金蓄，朱达荣，胡子粮，唐晟，雷经发
- **发表年份：** 2022
- **文献标题：** 基于 RGB 颜色空间的红外偏振人脸识别
- **期刊/会议名称：** 激光与光电子学进展
- **DOI/网址：** 3788/LOP202259.1233001

## 2. 研究背景与目的 (Why)

本研究聚焦于热红外成像在人脸识别中的问题（红外热像在夜间或弱光环境下具有广泛的应用价值），针对存在分辨率低、细节不清、边界模糊等不足，提出了一种人脸偏振热像 RGB 空间融合框架和一种基于支持向量机（SVM）的人脸识别方法（引入偏振成像技术并将红外偏振图像映射至 RGB 颜色空间），增强人脸红外热像的纹理和细节，提高人脸长波红外热像的结构相似性，使所提框架的精度达到 75.6%。

相关文献：

基于可见光人脸图像的识别准确率得到了很大提高，已经达到 99% 的识别率 [2-3]

Buddharaju 等 [8] 分析红外光谱提取的皮下解剖信息，进而提出一种基于血管网络特征的人脸识别方法，但是血管网络会受到高对比度区域（如发际线和皮肤）的影响，识别精度不高。

Xie 等 [9] 则根据 Pennes 方程通过对血液灌注的量化，将人脸表面的热信息转换为血液灌注率，再利用离散余弦变换域的可分性判别准则提取血液灌注信息的判别特征，提高了红外人脸识别系统的性能。

Hu 等 [10] 提取红外与可见光谱人脸图像的方向梯度直方图（HOG）等特征并利用二者的相似性和偏最小二乘法，提高了人脸红外热像识别的效果，识别率可达 49.9%。

Sarfraz 等 [11] 则通过深度感知映射将可见光谱人脸图像特征映射到红外光谱域，得到类红外热像人脸图像特征，然后对类红外热像人脸图像特征矢量与直接获取的红外热像人脸图像特征矢量之间的余弦相似性进行识别，提高了红外热像人脸识别的精度，在三个典型数据集上的测试精度可达 86%

美国陆军研究实验室（ARL）将偏振探测技术引入红外人脸热成像 [17]，有效凸显了红外人脸热像的细节特征，特别是在鼻子、嘴巴周围，补充了红外热成像所缺少的纹理细节

## 3. 研究方法与技术路线 (How)

本文提出一个**基于红外偏振信息与 RGB 色域融合的人脸识别框架**

人脸偏振成像：

在人脸热像检测的基础上融入偏振探测技术，利用IQU多维信息进行人脸表面的物理特性的反演与分析，降低测量噪声的影响，提高目标分割的精度，增加热图的区分度与识别率。

RGB 色域偏振热像融合：

将 I、Q、U 参量图像组合变换并映射到RGB 颜色空间进行融合，以此来突出图像的特征：对红外偏振图像进行均值滤波预处理，以防止噪声对图像带来的不利影响。

对滤波后的 Stokes 参量图像应用 DoG 处理（采用 DoG 边缘检测突出特征区域），突出人脸的具体细节并降低高频和低频噪声；之后再对处理过的图像进行权重处理，将 Stokes 参量 I、Q、U 图像对应色域通道 R、G、B，从而可以形成 RGB 色域偏振热像融合的彩色图像

红外偏振人脸识别算法：

使用HOG提取人脸特征，使用SVM分类器对训练样本类进行训练，获得 SVM 人脸识别模型对测试图像进行分类识别

数据集：自建实验平台采集红外偏振图像与可见光图像，建立数据集并几何归一化。

---

## 4. 核心内容与创新点 (What)

提出一种**基于 RGB 颜色空间的红外人脸热像色域融合框架**，并进行偏振热像人脸融合，形成 RGB 红外偏振融合图像：

在红外热成像的基础上，利用偏振成像获取人脸的红外热像（首次在红外人脸识别中引入偏振成像结合 RGB 融合方法）。

通过高斯差分（DoG）边缘特征自适应调节通道权重（构建“Stokes 参数 → DoG 边缘强化 → RGB映射”三步融合框架；）

一种**基于支持向量机（SVM）的红外热像人脸识别方法**（基于 HOG-SVM 的红外偏振人脸识别方法）：

利用HOG 提取 RGB 红外偏振融合图像的特征

利用红外偏振相机搭建了人脸红外偏振热像采集平台，模拟夜间环境，提高实验可信度；

---

## 5. 研究结论与实际意义

实验结果表明，所提方法的信息熵和结构相似性更高，并且得到的人脸特征更加接近可见光图像，人脸识别率更高。通过偏振探测获得的多通道图像（I、Q、U）能更好地凸显人脸关键区域（如眼睛、嘴巴、鼻子）的细节。将其融合为 RGB 图像后，结构相似性提升，信息熵更高，显著接近可见光图像的表现力。在 HOG + SVM 框架下，识别率比普通红外图像提高了近 20%。该方法对**夜间监控、身份验证与安防系统**等实际应用具有现实指导意义。

---

## 6. 本文的不足与未来展望

样本规模较小，仅限于室内实验

(RGB映射权重人为设定，未采用学习型融合方式；)

(未来可考虑：

- 引入 CNN 或 Transformer 对融合图像进行深层特征提取；
- 扩大数据集规模，并引入更多场景变化；
- 采用自适应融合策略或注意力机制优化 RGB 映射过程。)

## 7. 个人思考与选题关联

本文展示了**红外偏振信息**在提升图像可识别性方面的潜力，尤其是对于弱光环境人脸识别任务具有重要意义。我认为其 RGB 融合思路可与深度神经网络结合，进一步探索偏振信息在**边缘计算终端**（如夜间监控设备、安防机器人）中的部署策略，作为“**多模态弱光条件下人脸识别**”选题的参考。

“**多模态 (Multimodal)**”是人工智能领域中的一个核心概念，指的是**融合多种不同类型的数据模态**来进行感知、理解或生成。每种“模态”代表一种信息源，例如图像、文本、语音、视频、传感器数据等。

多模态 = 融合图像、文本、语音等多种数据类型来实现更全面的智能理解与交互。

模态	数据类型	示例
图像模态	图片/视频帧	看图识物、视觉描述
语言模态	文本/语音	文本理解、语音识别
音频模态	声波特征	语音、音乐、环境音
视频模态	图像 + 音频 + 时间信息	动作识别、视频分析
其他模态	传感器、触觉、脑电波等	医疗、自动驾驶、VR 等应用中使用

# 多模态系统典型任务

任务名称	说明
图文匹配 (Image-Text Matching)	给定一张图，找到最匹配的文字描述，或反过来
图像生成文本 (Image Captioning)	给图片自动生成一句话描述
文本生成图像 (Text-to-Image)	通过描述生成一张图，如 DALL·E
视觉问答 (Visual Question Answering, VQA)	给图片和一个问题，回答内容相关问题
多模态情感分析	综合语音+文字+表情来判断情感
多模态检索	输入文本找图，或输入图找视频等

## 应用场景非常广泛：

场景	多模态体现
智能客服	语音识别 + 文本理解 + 表情分析
自动驾驶	摄像头（图像）+ 雷达（点云）+ GPS
医疗影像	CT 图像 + 病人文本报告
教育辅助	视频讲解 + 语音 + 图文内容综合理解
跨模态检索	输入一段话，找到匹配的视频片段