摘要

玻璃等透明介质的界面反射现象因其物理特性呈现出难以避免的干扰特性。这类干扰不仅损害日常摄影的成像质量，还易在机器视觉主导的智能化场景中干扰感知，影响系统的决策可靠性。反射分离任务旨在从带反射干扰的混合图像（Mixture Image）中解耦玻璃两侧的场景信息，抑制玻璃反射并重建无干扰的视觉内容，从而获得背景图像（Transmission Image）和反射图像（Reflection Image）。针对反射分离领域对提升方法效果与降低数据获取难度的需求，本文研究基于分层先验引导实现反射分离，其中分层先验指通过相机特性、场景特性或用户交互提取的多维度显式约束，其能够提供反射或背景场景的局部或全局先验知识，以降低反射分离问题的病态性。本文着眼于现有方法依赖隐性约束、数据获取繁琐、缺少语义引导等三项研究挑战，从利用分层先验引入显性几何约束、减轻拍摄设备依赖、注入高层语义约束等三个维度的关键技术入手，提出了基于全景图像反射内容先验的反射分离方法、基于频闪光源时变光照先验的反射分离方法、基于语言描述场景语义先验的反射分离方法等三项方法，并结合真实反射数据验证方法的有效性。本文的主要创新点包括以下几个方面：

（1）针对现有研究依赖隐性约束的问题，借助全景图像相比于普通图像具有更广视场的优势，探索将全景图像中对应于玻璃反射的反射场景作为分层先验引入反射分离任务，利用反射内容先验提供的底层图像线索减轻问题病态性，解决了如何对齐几何与亮度层面的反射内容这一科学问题，为求解提供显性几何约束，实现高质量反射分离。为此，首先对反射场景与反射图像间的不对齐问题进行建模分析，并采用了两阶段全景图像反射分离方法 PIR^2（Panoramic Image Reflection Removal）在实现反射对齐后完成背景恢复。针对 PIR^2 用户交互不简便、反射对齐不稳定、数据合成不真实的问题，进一步提出了端到端全景图像反射分离方法 PAR^2Net（Panoramic Image Reflection Removal Network），通过单次用户交互完成待处理混合图像指定，利用自适应策略实现稳定反射对齐，改进数据合成方式减小合成与真实数据间差异。为验证方法有效性，构建了基于全景图像的真实反射分离数据集并将所提方法应用于其上，实验结果显示，基于反射内容先验引导的反射分离方法不仅在全景图像上相比于基于单张图像的方法效果明显提升，并且在没有全景相机的情况下也能够很好地适用于普通相机拍摄的视场有限的混合图像，具备良好的设备兼容性。

（2）针对现有研究数据获取繁琐的问题，发现频闪光源具有能够产生随时间快速周期变化光照的特性，借助这一特性克服对于特殊硬件设备的依赖，通过使用普通相机以短曝光时长拍摄图像序列记录频闪现象，将场景光照的时变特性作为线索，充分

利用图像层之间的光照差异，提取单侧频闪场景内容作为分层先验，解决了如何解耦运动与光照引起的强度变化这一科学问题，实现稳定的反射分离与频闪消除。为此，对同时存在反射干扰和频闪光照的图像成像模型进行了分析，通过频域分析与滤波对提取单侧频闪场景作为分层先验进行了可行性验证。对于包含动态内容的真实场景，提出了一个端到端频闪引导的反射分离方法 LIKENet（Light Flickering Guided Reflection Removal Network），利用对混合图像序列的短时和长时观测提取单侧频闪场景的上下文线索和非频闪部分的亮度一致线索，分别引导实现反射分离和闪烁消除。本文构建了一个包含频闪光照的合成和真实混合图像序列数据集，实验表明经时变光照先验引导的反射分离方法可以实现对于动态反射场景的稳定分离。此外，本文还展示了所提方法在视频频闪消除、高速脉冲相机反射分离等任务上的效果。

（3）针对现有研究缺少语义引导的问题，借助语言描述具有灵活指定图像层语义内容的优势，探索引入语言描述的多模态反射分离框架，利用用户指定的语言描述提供的场景语义作为分层先验，解决了如何建立内容混叠情况下的语义对应这一科学问题，通过注入高层语义约束显式引导混合图像中反射与背景的语义解耦，突破传统方法依赖单一图像模态的局限性，在无需硬件辅助的条件下，增强反射分离的可控性、场景适应性和模型的泛化能力。利用语言描述对场景内容的强指向性，设计自适应全局交互模块，联合学习图像特征与语言描述的跨模态语义一致性，利用语言门控机制确保语言描述与可分离图像层的对应关系，并通过随机训练策略进一步克服可识别层不确定的问题。为解决语言描述标注缺失的问题，提出基于配对图文对的数据合成方法，构建包含人工标注语言描述的反射分离数据集。实验表明，引入场景语义先验引导的反射分离方法在真实场景中能够在混合图像存在语义耦合的情况下有效分离反射与背景图像，为反射分离提供了兼具灵活性与鲁棒性的新范式。

（4）针对当前反射分离领域缺少切实可行的稳定落地方案的现状，引入预训练扩散模型的生成先验并融合语言描述场景语义先验，构建了基于语言描述引导的反射分离应用。开发了支持数据指定和结果对比的交互式应用，用户可通过语义引导强度和内容保真度参数实现反射分离过程的可控调节。通过将预训练扩散模型的图像生成能力与语言描述的场景语义约束进行多模态融合，突破传统卷积网络在强反射、过曝场景下的恢复能力局限，解决了复杂反射场景下反射分离难题。通过搭建反射采集装置获取真实反射图像，并结合互联网图像构建半合成数据集缓解纯合成数据与真实场景的数据分布差异。实验表明，所提方法在基准测试数据集上相比基于单张图像的方法性能显著提升，并在室内灯光、展览玻璃和飞机舷窗等高频日常场景实现高质量反射分离，为反射分离技术的实际应用提供了兼具生成质量与部署可行性的新方案。

关键词：反射分离，图像复原，深度学习

Decomposition prior guided reflection separation

Yuchen Hong (Computer Application Technology)
Supervised by Prof. Boxin Shi

ABSTRACT

As a core element in modern architecture and industrial design, semi-reflectors (such as glass) expand human living spaces while inevitably introducing visual perception interference due to its physical properties, *i.e.*, the reflection phenomena. This interference not only degrades the image quality in daily photography but also disrupts perception systems in machine vision-dominated intelligent scenarios, compromising decision-making reliability.

Layered priors refer to multi-dimensional explicit constraints extracted through camera characteristics, scene characteristics, or user interaction, which can provide local or global prior knowledge of the reflection or background scene to reduce the ill-posedness of the reflection separation problem. The reflection separation task aims to decouple scene information from both sides of glass surfaces in reflection-contaminated mixture images to obtain the transmission image and reflection image. To address current research demands for improving separation performance, reducing data acquisition complexity, and minimizing hardware dependencies, this thesis aims to achieve reflection separation with decomposition priors which refer to explicit constraints extracted through camera characteristics, scene characteristics, or user interaction, which can provide local or global prior knowledge of the reflection or transmission scene to reduce the ill-posedness of the reflection separation problem. This thesis focuses on three critical challenges in existing algorithms: Reliance on implicit constraints, cumbersome data acquisition processes, and insufficient semantic guidance. Through conducting explicit geometric constraints, reducing capture device dependencies, and injecting high-level semantic constraints, we propose three novel reflection separation algorithms: Panoramic image reflection separation, light flickering guided reflection separation, and language guided image reflection separation. Extensive experiments on real-world reflection data validate their effectiveness. The contributions of this thesis are summarized as follows:

(1) Leveraging the wider field-of-view advantage of panoramic images over conventional images, this thesis explores integrating reflection scenes corresponding to glass surfaces from panoramic imagery as decomposition priors into reflection separation tasks. By utilizing low-

level image cues provided by reflection content priors to mitigate the problem's ill-posedness, we address the scientific challenge of aligning reflection content at both geometric and photometric levels, thereby establishing explicit constraints to achieve high-quality reflection separation. We first analyze the misalignment problem between reflection scenes and reflection images, proposing a two-stage Panoramic Image Reflection Removal (PIR^2) algorithm that achieves transmission restoration after reflection alignment. PIR^2 extracts mixture images and reflection scenes from panoramic inputs through user interaction, implements coarse-to-fine reflection alignment via polynomial fitting-based data preprocessing, patch-matching algorithms, and a reflection refinement network, ultimately recovering the transmission image from the mixture image through a transmission restoration network using predicted reflection images. Addressing PIR^2's limitations in cumbersome user interaction, unstable alignment, and data discrepancy, we further propose an end-to-end Panoramic Image Reflection Removal Network (PAR^2Net). This enhanced framework enables single-step user interaction for mixture image specification, stable reflection alignment through adaptive strategies, and improved data synthesis pipelines that reduce the domain gap between synthetic and real data. To validate effectiveness, we construct a panoramic image-based real-world reflection separation dataset. Experimental results demonstrate that our decomposition prior-guided algorithms not only significantly outperform single-image baselines on panoramic inputs but also exhibit superior device compatibility. The proposed method also achieves robust performance on conventional camera-captured mixture images with limited FOVs without requiring specialized panoramic cameras.

(2) Leveraging the time-varying illumination characteristics of light flickering that produces rapid periodic intensity changes, this thesis eliminates existing algorithms' dependence on specialized hardware. By capturing image sequences using conventional cameras with shorter exposure times than standard photography to record flickering patterns, we exploit temporal illumination variations as critical cues. Through extracting unilateral flickering scene content as decomposition priors, we address the scientific challenge of decoupling intensity variations caused by motion versus illumination, achieving stable reflection separation and flicker removal. To this end, we first analyze the imaging model of images simultaneously contaminated by reflections and light flickering. Frequency-domain analysis and filtering validate the feasibility of extracting unilateral flickering scenes as decomposition priors. For real-world scenarios containing dynamic content, we propose LIKENet (LIght flicKEring guided reflection removal Network), an end-to-end framework that utilizes short-term observations

and long-term contextual cues from mixture image sequences to extract flickering context from unilateral scenes and brightness consistency in non-flickering components. These dual pathways respectively guide reflection separation and flicker removal. We construct comprehensive synthetic and real-world mixture image sequence datasets with light flickering. Experimental validation demonstrates that the proposed temporal illumination prior-guided algorithm achieves robust separation for dynamic reflection scenarios. Furthermore, we show the method's effectiveness in video flickering removal and high-speed camera reflection separation tasks.

(3) Capitalizing on the unique advantage of language descriptions in flexibly specifying semantic content for image layers, this thesis explores a language-visual collaborative reflection separation framework that leverages multi-modal information. By utilizing user-provided language descriptions as decomposition priors to convey scene semantics, we address the scientific challenge of establishing semantic correspondences under content aliasing conditions. The framework explicitly guides semantic decoupling between reflection images and transmission images in mixture images, overcoming traditional algorithms' limitations of relying solely on single modal image data. This breakthrough enhances separation controllability, scene adaptability, and model generalization capabilities without requiring multi-view inputs or specialized hardware assistance. Exploiting the strong scene-specific guidance of language descriptions, we design an Adaptive Global Interaction Module (AGIM) to jointly learn cross-modal semantic consistency between visual features and textual descriptions. A language gating mechanism ensures precise correspondence between language descriptions and separable image layers, while randomized training strategies further mitigate recognizable layer ambiguity. To resolve the lack of annotated language descriptions, we propose a paired text-image synthesis approach that constructs a reflection separation dataset with manually curated textual annotations. Experiments demonstrate that the scene semantic prior-guided reflection separation algorithm effectively disentangles reflection and transmission images under semantic coupling conditions in real-world mixture images, establishing a new paradigm that balances operational flexibility and separation robustness.

(4) To address the lack of practical and stable solutions for real-world deployment in reflection separation, this thesis introduces a language-guided reflection removal framework by integrating the generative prior of pre-trained diffusion models with semantic priors provided by language descriptions. We develop an interactive application supporting data specification and result comparison, enabling user-adjustable control through semantic guidance intensity

and content fidelity parameters. The proposed method overcomes the limitations of traditional convolutional networks in handling strong reflections and overexposed regions via multi-modal fusion of diffusion models' image generation capabilities and language-based semantic constraints. A semi-synthetic dataset is constructed using real reflection images captured through custom acquisition devices combined with internet-sourced images, effectively mitigating the distribution gap between synthetic and real-world data. Experimental results demonstrate that the solution achieves superior performance compared to single-image baselines and enables high-quality reflection separation in challenging real-world scenarios including indoor lighting conditions, exhibition glass cases, and aircraft window reflections, providing a practical framework that balances restoration quality with deployment feasibility.

KEY WORDS: Reflection separation, Image restoration, Deep learning