摘要

动态场景清晰成像是计算摄像与计算机视觉中的重要研究方向，其目标是在运动、景深变化等复杂条件下获得结构清晰、焦点可控的图像或视频。计算成像技术通过结合传感器与计算算法突破传统光学成像的限制，例如在动态运动场景中消除模糊，在浅景深条件下扩展景深并恢复全聚焦图像，以及在动态场景中支持后捕获重对焦，在自动驾驶、机器人感知、工业检测以及科学成像等领域具有广泛需求。

传统相机的快门速度与光圈大小之间存在固有耦合约束：延长曝光会引入运动模糊，缩短曝光并增大光圈则压缩景深、导致失焦模糊。两类退化在动态场景中往往同时出现并相互耦合，使得仅依赖单帧或低帧率二维图像进行恢复成为高度不适定问题，即解非唯一且对观测噪声和先验假设高度敏感的逆问题。仅基于传统图像的恢复方法虽然可以借助图像先验与深度神经网络取得一定效果，但其观测维度始终受限于曝光后的模糊图像，难以稳定恢复时间维和深度维同时丢失的信息，尤其在空间变化运动、遮挡及弱纹理区域表现不佳。

以事件相机和脉冲相机为代表的神经形态相机凭借微秒级时间分辨率和异步响应特性，能够记录传统成像过程中丢失的高频时间变化信息，为动态场景清晰成像提供了新的观测维度。本文分别从运动模糊消除、景深扩展与全聚焦成像、动态重对焦三个方面切入，研究如何将神经形态信号转化为面向具体成像任务的有效约束：（1）事件流异步稀疏且单色事件与彩色图像存在模态差异，如何实现鲁棒表征与跨模态融合是运动去模糊的基本难点；（2）传统焦点堆栈依赖多次静态曝光，如何利用神经形态信号实现单次扫描下的连续深度采样尚属空白；（3）动态重对焦中失焦效应与运动效应在同一异步流中耦合，如何基于编码焦点堆栈实现二者解耦并在时间和焦点两个维度上联合恢复未被解决。

（1）针对运动去模糊中事件流异步稀疏、单色事件与彩色图像存在明显模态差异的问题，提出基于神经事件帧表征与跨模态分解的运动模糊去除方法。该方法通过双向时序建模与可学习的空间特征提取，将原始事件流表征为鲁棒的稠密表示，使运动轨迹和时间关系能够稳定地进入去模糊过程，并构建事件引导的去模糊框架以恢复空间变化运动模糊。针对单色事件与彩色图像之间的模态差异，将彩色图像分解为亮度分量与颜色分量，在亮度域完成事件融合后再引导颜色补偿，以减少颜色溢出与边界染色。同时构建真实彩色事件去模糊数据集。实验表明，所提方法在仿真与真实数据集上相对现有方法在客观指标和视觉质量上均有提升。

（2）针对浅景深条件下单图失焦恢复观测不足、传统焦点堆栈依赖多次静态曝光

的问题，提出基于神经形态焦点堆栈的景深扩展方法。该方法在单次连续焦点扫描中同步记录事件或脉冲信号，以高时间分辨率记录深度维度上的焦点变化，既保留了焦点堆栈在不同深度可分离信息的优势，又实现了连续而非离散的深度采样。在此基础上设计由重对焦网络和堆栈融合网络组成的统一数据驱动框架，分别用于焦点堆栈重建与像素级融合权重预测，支持任意数量重对焦图像输入。实验表明，所提方法在严重失焦、遮挡和有限运动条件下，相对单图去失焦、双像素和传统焦点堆栈方法取得更稳定的全聚焦结果，并验证了该方法对事件相机与脉冲相机的通用性。

（3）针对动态重对焦中失焦效应与运动效应在同一异步流中耦合、难以同时恢复目标时间和目标焦点的问题，提出基于编码事件焦点堆栈的失焦-运动效应解耦方法。该方法通过周期性调制焦点扫描，使扫描触发事件在时间上呈现近似对称结构，为运动事件与扫描事件的分离提供观测基础。两阶段框架中，第一阶段利用运动事件恢复目标时刻的中间帧，第二阶段利用扫描事件将中间帧重对焦到任意目标焦点，并结合基于选定区域的迭代搜索策略实现后捕获焦点追踪，可在拍摄完成后对任意目标区域进行自动对焦。实验表明，所提方法在仿真与真实动态场景上，相对单图重对焦与视频重对焦方法取得更稳定的动态重对焦和焦点追踪结果。

本文分别从运动模糊消除、景深扩展与全聚焦成像、动态重对焦三个方面切入，提出了基于神经形态显式约束的动态场景清晰成像方法，为神经形态相机在计算成像领域的应用提供了新的技术方案。

关键词：动态场景清晰成像，神经形态相机，模糊消除

Neuromorphic Camera Guided Dynamic Scene Imaging

Minggui Teng (Computer Application Technology)
Supervised by Prof. Boxin Shi

ABSTRACT

Imaging in dynamic scenes is an important research direction in computational photography and computer vision. Its goal is to acquire images or videos with clear structure and controllable focus under complex conditions such as motion and depth-of-field variation. By combining sensors with computational algorithms, computational imaging techniques overcome the limitations of conventional optical imaging, including removing blur in dynamic motion scenes, extending the depth of field and recovering all-in-focus images under shallow depth of field, and supporting post-capture refocusing in dynamic scenes. These capabilities are widely needed in autonomous driving, robotic perception, industrial inspection, and scientific imaging.

Conventional cameras are subject to an inherent coupling constraint between shutter speed and aperture size: extending the exposure introduces motion blur, while shortening the exposure and enlarging the aperture compress the depth of field and cause defocus blur. These two degradations often coexist and interact in dynamic scenes, making restoration from single-frame or low-frame-rate two-dimensional images a highly ill-posed inverse problem, where the solution is non-unique and highly sensitive to observation noise and prior assumptions. Although conventional image-based restoration methods can exploit image priors and deep neural networks to achieve certain results, their observations are inherently limited to post-exposure blurry images. It is therefore difficult for them to stably recover information that is simultaneously lost along the temporal and depth dimensions, especially in regions with spatially varying motion, occlusion, and weak texture.

Neuromorphic cameras, represented by event cameras and spike cameras, can record high-frequency temporal variations lost in conventional imaging with microsecond-level temporal resolution and asynchronous response, providing a new observation dimension for clear imaging in dynamic scenes. This thesis investigates how to transform neuromorphic signals into effective constraints for specific imaging tasks from three aspects: motion blur removal, depth-of-field extension and all-in-focus imaging, and dynamic refocusing. The main chal-

lenges are as follows: (1) event streams are asynchronous and sparse, and monochromatic events have a modality gap with color images, making robust representation and cross-modal fusion a fundamental difficulty in motion deblurring; (2) conventional focal stacks rely on multiple static exposures, and how to use neuromorphic signals to achieve continuous depth sampling in a single sweep remains unexplored; (3) defocus and motion effects are coupled in the same asynchronous stream during dynamic refocusing, and how to disentangle them based on coded focal stacks and jointly recover the temporal and focal dimensions remains unresolved.

(1) To address the asynchrony and sparsity of event streams and the pronounced modality gap between monochromatic events and color images in motion deblurring, this thesis proposes a motion blur removal method based on neural event frame representation and cross-modal decomposition. Through bidirectional temporal modeling and learnable spatial feature extraction, raw event streams are represented as a robust dense representation, allowing motion trajectories and temporal relationships to be stably incorporated into the deblurring process. An event-guided deblurring framework is then constructed to recover spatially varying motion blur. To address the modal gap between monochromatic events and color images, the color image is decomposed into luminance and color components. Event fusion is performed in the luminance domain and then used to guide color compensation, reducing color bleeding and boundary staining. A real-world color event deblurring dataset is also constructed. Experiments show that the proposed method improves upon existing methods on synthetic and real datasets in objective metrics and visual quality.

(2) To address the insufficient observations available for single-image defocus restoration under shallow depth of field and the reliance of conventional focal stacks on multiple static exposures, this thesis proposes a depth-of-field extension method based on neuromorphic focal stacks. The method synchronously records event or spike signals during a single continuous focal sweep, capturing focal changes along the depth dimension with high temporal resolution. It retains the advantage of focal stacks in separating information at different depths while enabling continuous rather than discrete depth sampling. Based on this method, a unified data-driven framework composed of a refocusing network and a stack fusion network is designed for focal stack reconstruction and pixel-wise fusion weight prediction, respectively, and supports an arbitrary number of refocused image inputs. Experiments show that under severe defocus, occlusion, and limited motion, the proposed method achieves more stable all-in-focus results than single-image defocus deblurring, dual-pixel, and conventional focal stack methods, and

the generality of the method is validated on both event cameras and spike cameras.

(3) To address the coupling between defocus and motion effects in the same asynchronous stream during dynamic refocusing, where the target time and target focus are difficult to recover simultaneously, this thesis proposes a defocus-motion effect disentanglement method based on coded event focal stacks. By periodically modulating the focal sweep, the method makes sweep-triggered events exhibit an approximately symmetric temporal structure, providing an observation basis for separating motion events from sweep events. In the two-stage framework, the first stage uses motion events to recover an intermediate frame at the target time, while the second stage uses sweep events to refocus the intermediate frame to an arbitrary target focus. Combined with an iterative search strategy based on a selected region, the method further enables post-capture focus tracking, allowing automatic focusing on any target region after capture. Experiments show that on synthetic and real dynamic scenes, the proposed method achieves more stable dynamic refocusing and focus tracking results than single-image refocusing and video refocusing methods.

This thesis studies imaging in dynamic scenes from three aspects: motion blur removal, depth-of-field extension and all-in-focus imaging, and dynamic refocusing. It proposes methods based on explicit neuromorphic constraints, providing new technical solutions for applying neuromorphic cameras to computational imaging.

KEY WORDS: Dynamic Scene Imaging, Neuromorphic Camera, Blur Removal