中文摘要：

传统相机经过数十年的发展已经深入到现代社会生产生活的各个角落，相机成像的实用价值已经从满足人们摄影需求扩展到了赋能人脸识别、自动驾驶等信息时代的重要应用领域。然而，传统相机常常因为帧率不够快、动态范围不够高、功率过高、数据存储容量过大等因素而影响摄影体验以及在下游任务上的表现。近年来，包括事件相机和脉冲相机在内的神经形态相机作为一类新型的传感器引起了越来越多研究者的关注。神经形态相机借鉴了生物视网膜的成像原理，采用异步动态感知的方法，高时间精度且宽动态范围地获取场景内的光强信号。其中事件相机的像素异步感知场景光强的变化，而脉冲相机的像素异步记录光强的积分。然而，由于受到相机制造工艺水平和设计成本等因素的制约，当前主流的神经形态相机存在分辨率低、噪声大等问题，这制约了神经形态相机在实际应用场景中的广泛使用。相较于传统相机对应的图像处理领域已拥有成熟的图像信号处理算法，尚缺少针对神经形态相机所拍神经形态信号进行有效去噪和超分辨处理的算法。针对这一棘手问题，本文研究神经形态信号的工作原理，提出了两种去噪和超分辨率处理算法，且通过实拍数据验证了算法的有效性。针对神经形态信号本身的研究充分证明了高质量的神经形态信号在辅助传统图像成像方面的重要意义和巨大潜力，因此，本文进一步提出在混合相机系统中利用神经形态事件相机辅助矫正卷帘快门图像的算法。本文在原理、数据、方法方面的创新具体包括以下几个方面：

（1）针对传统相机和神经形态事件相机在分辨率、响应时延、噪声程度等方面存在的优势互补，本文探索将具有高空间分辨率和低传感器噪声特性的传统基于帧的传感器和具有高速和高动态范围特性的新兴事件传感器相结合进行混合成像，通过联合滤波来实现优势互补，从而实现高性能成像。为此，本文提出将传统帧信号和事件信号进行信息关联的理论模型，并采用联合引导滤波的方法（Guided Event Filtering, GEF）来提升事件相机输出事件信号的质量。GEF 首先根据光流模型对原始事件进行运动补偿，将原始事件补偿到引导图像的平面上，并计算出对应图像的时间梯度图。然后对两者执行联合滤波，通过联合滤波辅助事件帧从图像梯度图中获取精细的纹理结构，同时使事件帧实现噪声的消除和分辨率的提升。当引导图片表现不佳时，GEF 会采用事件自引导机制，该机制会借助邻近事件进行引导滤波。为了验证算法的有效性，本文使用事件相机和传统相机构建了混合相机系统，并将所提算法应用于真实数据，实验结果显示经过引导滤波之后的事件实现了明显的分辨率提升和噪声消除。此外，本文还展示了去噪超分辨率后的事件信号在高帧率视频合成、高动态范围成像、角点检测跟踪等方面带来的性能提升。

（2）针对 GEF 算法依赖引导图像的质量，且信号处理速度难以满足现实需求的问题，本文进一步提出了基于真实样本学习的神经形态信号去噪和超分辨率算法，称为 NeuroZoom。NeuroZoom 使用 3D U-Net 作为主干架构，利用 3D 卷积网络来提取神经形态信号中丰富的时域关联特征，并利用金字塔型网络结构和多尺度同步监督等机制来实现不同分辨率的神经形态信号去噪和超分辨率。为学习低分辨率和高分辨率神经形态信号之间的对应关系，本文提出了一种基于真实拍摄的用于多分辨率神经形态信号数据收集的“显示器-相机”系统。该系统通过拍摄显示器中播放的高帧率视频来收集三个尺度的神经形态信号数据。NeuroZoom 借鉴 Noise-to-Noise 的训练模式，使用带有同分布噪声的“高分辨率—低分辨率”成对数据集进行网络训练，进而有效地执行高达 4 倍的神经形态事件信号去噪、超分辨率任务，和 2 倍的神经形态脉冲信号去噪和超分辨率任务。验证实验证明 NeuroZoom 可以快速实现有效的神经形态信号处理，且增强的信号有助于提高在视觉对象跟踪和图像超分辨率重建等任务上的性能。

(3) 针对 CMOS 传感器中采用卷帘快门工作机制导致的拍摄运动场景时图像中会产生边缘扭曲和区域遮挡的问题，即卷帘快门效应，本文提出利用神经形态事件信号矫正卷帘快

门图像，将一张卷帘快门图像恢复成一段高帧率全局快门视频的算法EvUnroll。EvUnroll引入了一种新颖的计算摄像系统，该系统包含一个卷帘快门传统相机和神经形态事件相机，并针对该问题设计了一个包含四个模块的神经网络模型。该算法通过事件信号建立起了卷帘快门图像和全局快门图像之间的时空关联，提出了分别用于解决卷帘快门图像中存在的边缘扭曲和区域遮挡问题的光流估计模块和残差估计模块。两个模块的中间结果通过一个融合模块得到最后矫正后的图像。此外，针对输入图像中可能存在运动模糊的问题，该网络还包含了一个可选的去模糊预处理模块。在验证测试中，本文利用帧率达5700FPS的高速相机拍摄得到的高速视频获取仿真数据集并用于网络训练，同时搭建了一套协同卷帘快门相机和事件相机的混合相机系统，用于拍摄测试所需的真实数据。实验结果证明所提出的方法在仿真数据集和真实数据集上都取得了显著优于当前最优方法的性能表现。

English abstract:

After decades of development, traditional cameras have penetrated all corners of production and life in modern society. The practical value of camera imaging has expanded from meeting people's photography needs to empowering important application fields in the information age such as face recognition and autonomous driving. However, traditional cameras often affect the photography experience and the performance of downstream tasks due to factors such as insufficient frame rate, insufficient dynamic range, and high power. At present, many vision and robotics tasks rely on the powerful processing performance of imaging equipment for high-speed motion and high dynamic range (HDR) scenes, and can meet the characteristics of high spatial resolution and low noise. In recent years, neuromorphic cameras have attracted the attention of more and more researchers as a new type of imaging sensor. The neuromorphic camera draws on the imaging principle of the biological imaging system, adopts the method of asynchronous dynamic sensing, and obtains the light intensity change signal in the scene with high time accuracy and wide dynamic range. However, due to the constraints of camera manufacturing process level and design cost, the current mainstream neuromorphic cameras have the problems of low resolution and large noise, which will restrict the wide use of neuromorphic cameras in practical application scenarios. However, there is currently no algorithm for effective denoising and super-resolution processing of neuromorphic signals captured by neuromorphic cameras. In response to this problem, this thesis studies the imaging principle of neuromorphic signals, and proposes two contributions, and verifies the effectiveness of the algorithm through real shot data. Studies on neuromorphic signals themselves fully prove the importance and great potential of high-quality neuromorphic signals in assisting traditional imaging. Therefore, this paper further proposes an algorithm to use neuromorphic event cameras in hybrid camera systems to assist in the correction of rolling shutter images. The main contributions of this thesis include:

(1) This chapter proposes high-performance imaging by exploring the synergy between traditional frame-based sensors with high spatial resolution and low sensor noise, and emerging event-based sensors with high speed and high dynamic range. We introduce a novel computational framework, termed Guided Event Filtering (GEF), to process these two streams of input data and output a stream of super-resolved yet noise-reduced events. To generate high-quality events, GEF first registers the captured noisy events onto the guidance

image plane according to our flow model. It then performs joint image filtering that inherits the mutual structure from both inputs. Lastly, GEF re-distributes the filtered event frame in the space-time volume while preserving the statistical characteristics of the original events. When the guidance images underperform, GEF incorporates an event self-guiding mechanism that resorts to neighbor events for guidance. We demonstrate the benefits of GEF by applying the output high-quality events to existing event-based algorithms across diverse application categories, including high-speed object tracking, depth estimation, high frame-rate video synthesis, and super-resolution/HDR/color image restoration.

(2) This chapter proposes a neuromorphic signal denoising and super-resolution algorithm, called NeuroZoom, based on real sample learning, to address the problem that the performance of the GEF algorithm depends on the quality of the guiding image, and the signal processing speed cannot meet the requirements of practical applications. NeuroZoom uses 3D U-Net as the backbone architecture to extract rich temporal correlation features from neuromorphic signals. It also employs a pyramid network structure and multi-scale supervision mechanisms to achieve neuromorphic signal denoising and super-resolution at different resolutions. To establish the correspondence between low-resolution (LR) and high-resolution (HR) neuromorphic signals, a ``display-camera'' system for collecting multi-resolution neuromorphic signal data is proposed, which is based on real shooting. This system collects neuromorphic signal data of three scales by capturing high-frame-rate videos played on a fixed display. NeuroZoom adopts the noise-to-noise training mode, using a paired data set of ``HR-LR'' with the same distribution noise to train the network, and thus effectively performs up to 4$\times$ neuromorphic event signal denoising and super-resolution tasks, and 2$\times$ neuromorphic spike signal denoising and super-resolution tasks. Experimental results demonstrate that NeuroZoom can quickly and effectively process neuromorphic signals, and the enhanced data helps improve performance in tasks such as visual object tracking and image super-resolution reconstruction.

(3) This chapter proposes the algorithm EvUnroll for correcting rolling shutter images using neuromorphic event signals to address the problems of edge distortion and region occlusion caused by the rolling shutter effect that is prevalent in CMOS camera sensors. EvUnroll introduces a novel computational imaging system that includes both a rolling shutter traditional camera and a neuromorphic event camera, and designs a neural network model consisting of four modules. The algorithm establishes the spatiotemporal correlation between the rolling shutter image and the global shutter video through event signals and proposes optical flow estimation and residual estimation modules to solve the problems of edge distortion and region occlusion in the rolling shutter image, respectively. The intermediate results of the above two modules are fused to obtain the final corrected image. In addition, the network also includes an optional deblurring preprocessing module to address the problem of motion blur in the input image. In the validation test, we use a high-speed camera with a frame rate of $5700FPS$ to obtain simulated data for network training and build a hybrid camera system consisting of a rolling shutter camera and a collaborative event camera to capture real data for testing. Experimental results demonstrate that the proposed method achieves significantly better performance on both simulated and real data than the state-of-the-art methods.