摘要

视觉是高级生物获取外界环境信息最重要的方式。近年来随着集成电路、计算机与通信等领域的飞速发展，人工智能领域内对计算机视觉的研究也蓬勃发展起来。在诸如自动驾驶、无人机与科学观测等领域，新兴的应用对高速机器视觉提出了更高的要求，传统数字相机难以满足对高帧率的需求。近年来，研究者们提出了一类超高速的仿生相机，也即神经形态相机，这类相机抛弃了传统相机单次曝光成像的模式，仿照生物视网膜的成像机制进行各像元的异步工作。神经形态相机主要包括仿照视网膜外周进行差分型采样的事件相机与仿照视网膜中央凹进行积分型采样的脉冲相机。本文主要围绕其中的脉冲相机开展研究，其各像元对光子进行连续累积，并当累积值达到发放阈值时就立即发放一个二值脉冲并重置积分值。通过上述“积分-发放”的成像机制，脉冲相机可以二值脉冲流实现对场景以极细的时间粒度进行记录。

由于脉冲相机对场景记录的连续性，其在高速场景的清晰成像中展现出了很大的潜力。然而，从脉冲流中重建清晰影像也存在若干挑战。在脉冲流中，单个二值脉冲蕴含的信息十分有限，所以对于重建清晰的影像而言，利用一定时间内的脉冲流是必要的。而在脉冲相机对场景进行连续记录的过程中，场景中的运动也被记录在了脉冲流中，如果要利用不同时刻的脉冲，运动对齐是一个关键步骤。所以，对脉冲流中所蕴含运动信息的分析也是脉冲相机相关的重要研究主题。该任务一般基于不同时刻之间的像素的相关与匹配实现，而由于脉冲在记录过程中受到光子到达的泊松效应、电路的暗电流噪声、时钟控制脉冲读出导致的量化效应等影响，脉冲流中存在随机波动性，这给稳定的光强特征的提取带来了困难。上述脉冲的随机波动性同样也会给脉冲相机的影像重建带来挑战。综上脉冲相机的影像重建和运动分析是两个十分相关的基础问题，二者共同面临从脉冲流中获得稳定光强信息的挑战，且运动分析是影像重建的重要基础。本文即围绕这两个重要的问题，开展了面向高速脉冲数据的运动分析与影像重建，本文的主要贡献点包括：

1. **提出了基于层次化时空融合表征的脉冲光流估计方法**

   光流估计是运动分析的重要手段之一，而构建准确的相关张量是实现高质量光流估计的关键步骤。然而，光子到达的泊松噪声、电路热噪声以及时钟控制脉冲读出导致的量化噪声，使得脉冲流中提取的特征包含随机波动性，这会给构建的相关张量中引入歧义与模糊，从而影响最终光流估计的结果。为了抑制脉冲中波动性与随机性对光流估计的影响，本文提出了层次化的时空融合（Hierarchical Spatial-Temporal, HiST）表征来增强网络对脉冲的特征提取能力。HiST 针对多个时刻提取特征，并对不同时刻的特征进行层次化的时空融合。基于 HiST 表征，

本文提出了面向脉冲的光流估计网络 HiST-SFlow。实验结果表明，相比现有方法，所提的方法在合成数据与真实数据上都能实现较为准确的光流估计。

2. **提出了基于多相关张量联合解码的连续脉冲光流联合估计方法**

现有的运动分析多基于两个时刻点之间的光流估计实现。对于脉冲相机而言，其对场景的记录在时域上具有很强的连续性。这种连续性给运动分析的提升带来了更多的灵活性。本文将两时刻点之间的光流估计拓展到了多时刻点之间的光流联合估计。为了实现稳定的光强特征提取与相关张量构建，本文首先提出了脉冲发放时间差分的概念，并在此基础上提出了对偶脉冲发放时间差分表征，用于从脉冲流中提取稳定的光强信息。为了利用运动连续与脉冲连续的一致性，本文提出了多相关张量联合解码模块，将不同阶段的光流互相作为彼此的上下文信息加以利用，并且还在此基础上提出了全局运动信息仓库聚合，将所有运动的特征编码为一个信息仓库特征，在每段光流的解码时从中自适应地挖掘上下文信息。此外，为了实现连续光流的训练与评估，本文提出了一个基于真实场景的脉冲光流数据集。实验结果表明，所提方法在合成数据、实拍数据与下游任务上都能取得较优的性能。

3. **提出了基于多粒度对齐的脉冲相机影像重建方法**

在连续的脉冲流中，单个二值脉冲所蕴含的信息十分有限，重建清晰的影像需要利用一定时间内的脉冲流。而在脉冲对场景连续记录的过程中，场景中的运动也被记录在了脉冲流中，所以在融合不同时刻脉冲流的信息时需要进行运动对齐。而由于脉冲流中存在多种因素引起的随机波动性，脉冲流的光强特征提取会受到影响，进而影响运动的对齐。本文首先分析了脉冲流中时钟控制读出引起的量化效应的统计特性，进而提出了用于提取稳定光强特征的多阶脉冲发放时间差分融合表征。考虑到不同时刻的脉冲特征中会留存有脉冲流波动性带来的影响，本文提出了多粒度特征对齐模块，其中粗粒度的对齐由基于搜索策略的块级交叉注意力进行，细粒度的对齐由可变形卷积进行。基于上述针对脉冲流波动性的设计，本文所设计的神经网络相较于现有神经网络在合成数据与实拍数据上都取得了较为明显的性能提升。

关键词：脉冲相机，神经形态视觉，影像重建，运动分析，光流估计

Research on Motion Analysis and Image Reconstruction for Spike Cameras

Rui Zhao (Computer Applied Technology)
Supervised by Prof. Ruiqin Xiong

ABSTRACT

Vision serves as the most critical pathway for advanced organisms to acquire information from environments. With rapid advancements in integrated circuits, computer science, and communication technologies, research in computer vision has flourished in the artificial intelligence area. Emerging applications such as autonomous driving, unmanned aerial vehicles, and scientific observation demand high-speed machine vision capabilities that exceed the limitations of conventional digital cameras in achieving ultra-high frame rates. In recent years, neuromorphic cameras, a bio-inspired ultra-high-speed imaging devices, have been proposed. These cameras abandon the traditional single-exposure imaging mechanism and mimic the asynchronous working mechanism of retina. Neuromorphic cameras include event camera and spike camera. Event camera emulates peripheral retinal and works with a differential sampling model, while spike camera emulates the fovea and works with a intergral sampling model. This thesis focuses on spike camera. Each pixel of the spike camera continuously accumulates photons. Whenever the accmulation reaches a predefined threshold, a binary spike is fired and the accmulation is reset to zero. Through this ”integrate-and-fire” mechanism, spike camera achieves scene recording of ultra-fine temporal resolution.

The continuous recording capability of spike cameras demonstrates significant potential for high-speed motion imaging. However, reconstructing clear images from spike streams faces multiple challenges. Individual binary spikes contain limited information, necessitating the utilization of temporal spike sequences for reconstruction. However, during continuous recording for scenes, motion in scenes is also recorded in spike streams, and we need motion alignment when fusing spike streams in a temporal window. Consequently, motion analysis in spike streams emerges as a critical research topic. Most motion analysis methods rely on pixel-level correlation and matching across temporal frames, but spike streams have inherent random fluctuations casued by multiple factors. These factors include Poisson-distributed photon arrival, circuit dark current noise, and clock-controlled quantization effects. The fluctuations

complicate the extraction of stable intensity features, and they also pose challenges for image reconstruction. Thus, image reconstruction and motion analysis in spike cameras are fundamentally intertwined, both requiring robust intensity feature extraction, with motion analysis serving as the foundation for reconstruction. This thesis addresses these dual challenges, with key contributions as follows:

1. **Hierarchical spatial-temporal fusion for spike camera optical flow estimation**
   Optical flow estimation is one of the pivotal approaches for motion analysis, where constructing accurate correlation volumes is essential. However, random fluctuations in spike streams introduce ambiguity into correlation volumes, degrading optical flow performance. To mitigate this, we propose a Hierarchical spatial-Temporal Fusion (HiST) representation that enhances spike stream features through multi-temporal feature extraction and hierarchical fusion. Based on HiST, we develop HiST-SFlow, a neural network for spike camera optical flow estimation. Experimental results demonstrate superior accuracy on both synthetic and real-world datasets compared with state-of-the-art methods.
2. **Joint estimation of multiple motion fields for spike camera**
   Existing motion analysis frameworks typically focus on pairwise optical flow, limiting motion analysis for spike camera. Leveraging the temporal continuity of spike streams, we extend pairwise estimation to joint estimation for multiple motion fields. First, we propose the differential of spike firing times (DSFT), and propose a dual DSFT representation to extract stable light-intensity information. To exploit motion continuity, we design a joint correlation decoding module that use multiple motion feilds as contextual information for each other. Besides, we propose a global motion bank aggregation module for adaptive feature aggregation among differnt motion feilds. Additionally, we construct a dataset for based on real-captured data for joint estimation of multiple motion feilds for spike camera. Experimental results demostrate the effectiveness of the proposed method on synthetic data, real-world captures, and downstream tasks.
3. **Spike camera image reconstruction based on multi-granularity alignment**
   Reconstructing images from spike camera requires aggregating information in a temporal window, which inherently involves motion alignment. To address feature instability caused by random fluctuations, we first analyze the statistical properties of quantization effects caused by clock-driven spike readout and propose a multi-order DSFT (diffrential of spike firing time) fusion representation. We further develop a multi-granularity alignment module, where the coarse alignment is based on patch-level cross-attention

with a seaching strategy, and the fine alignment is based on deformable convolutions. Our network achieves significant performance gains over existing methods on both synthetic and real-world benchmarks.

KEY WORDS: Spike camera, neuromorphic vision, image reconstruction, motion analysis, optical flow