摘要

神经形态传感器是近年来新兴的视觉传感器，其通过对场景光强信息进行连续采样，输出高时间分辨率的异步神经形态信号（脉冲或事件）。相对于传统视觉传感器，神经形态视觉传感器具有高动态范围、高时域分辨率、低功耗的优势。在自动驾驶、无人机视觉导航、工业检测及视频监控等机器视觉领域，尤其在涉及高速运动或极端光照场景下，神经形态视觉传感器有着巨大的潜力。然而，与传统基于帧的“所见即所得”的成像范式不同，新型的异步神经形态信号无法直观地被人们所理解。因此，如何根据神经形态信号连续采样的特性，从稀疏的神经形态数据流中重建出准确的连续视觉图像，是神经形态视觉传感器亟需解决的基本问题。

本文首先形式化定义了神经形态视觉重建问题，给出了评测基准，并在此基础上从不同层次对连续视觉重建开展了系统性的研究：在数据层次，研究了从事件相机、脉冲相机的单模态神经形态数据，到事件数据和脉冲数据联合的多模态连续视觉重建；在算法层次，围绕生物合理性和计算有效性的原则，从手工设计的浅层脉冲神经网络出发，研究了数据驱动的深层脉冲神经网络和人工神经网络。本文的主要创新工作包括：

第一，针对脉冲数据的时空特征自适应提取问题，提出了一种基于生物脉冲处理机制的时空自适应脉冲神经网络算法，能有效实现异步的脉冲相机连续视觉重建。该算法首先根据脉冲数据的时域特性定义了运动置信矩阵，将时空动态神经元提取最优化问题建模为一阶马尔科夫随机场问题并求解。基于动态神经元提取的结果，该算法设计了一种自适应脉冲时序调整方法，通过引入生物启发的积分泄漏发放神经元以及基于突触时间相关可塑性规则的突触连接驱动，实现了视觉纹理的自适应连续重建。为测试算法的有效性，本文构建了首个用于脉冲相机视觉重建的数据集，即 PKU-Spike-Recon 数据集。实验结果表明，该算法在普通和高速场景下都能获得更好的主观评测质量及更高的二维熵和锐利度评价指标，并具有较高的鲁棒性。

第二，针对神经形态数据的深层特征提取问题，提出了采用自适应膜电位神经元的深层循环脉冲神经网络算法，能有效实现低功耗的脉冲和事件相机连续视觉重建。该算法构建了基于积分泄漏发放神经元和膜电位神经元的脉冲神经网络框架，对脉冲或事件数据进行有监督训练，首次完成了基于深层脉冲神经网络的大规模图像回归任务。网络内部的信息由二值脉冲传递，通过脉冲跳层连接、脉冲多尺度特征提取等方式提取深层脉冲特征，并最终通过膜电位神经元把二值脉冲信息转化为连续的灰度值输出，得到连续视觉重建结果。为进一步提升网络的时域感受野，本文还提出了膜电位辅助的混合 SNN-ANN 框架。实验结果表明，该算法在室内复杂光照和室外自动驾驶等多

种场景的性能与基于人工神经网络的模型相当，所提出两种框架的能耗分别比其对应的人工神经网络结构低 19.36 倍和 7.75 倍。

第三，针对多模态神经形态数据的特征融合问题，提出了一种脉冲和事件数据流多尺度融合的视觉重建网络算法，能有效实现复杂光照环境下的神经形态相机连续视觉重建。从仿生的角度来看，事件相机和脉冲相机分别模拟了视网膜外周和中央凹，因此通过融合这两种数据可以达到同时感知高动态范围的运动信息和精细的纹理信息的目的。针对脉冲和事件数据，该算法提出了基于自适应时域信息提取机制的神经形态数据表征方式，设计了基于编解码结构的网络对两类神经形态数据的特征进行提取，并通过引入特征融合模块有效刻画了两者的互补特性。为验证该算法的有效性，本文搭建了联合脉冲和事件相机的采样装置，并构建了 PKU-Retina-Recon 多模态脉冲数据集。实验结果表明，该算法能够有效地结合脉冲和事件数据重建出高动态范围的信息，取得更好的主观评测结果；在高速、低光照测试场景下的 SSIM 指标比单事件数据和单脉冲数据的重建结果分别提高 21% 和 16%。

第四，构建了面向高速运动对象的神经形态连续视觉重建系统，使用 FPGA 和上位机结合的方式实现了神经形态视觉重建算法的真实场景验证。传统的高速成像系统大多采用图像帧范式成像体制，通过“稠密”成像直接得到图像帧序列，会导致“海量数据”和“带宽有限”的矛盾。本文构建的神经形态视觉重建系统则以时空脉冲阵列信号的形式来高效表示与处理连续视觉信息，具有高时域分辨率、数据冗余少和低功耗等优势。具体地，该系统把所提浅层脉冲神经网络的动态神经元提取和自适应阈值重建进行了简化，使其可以部署在 FPGA 上并实时运行；为获得更高质量的重建结果，还设计了两种深层脉冲神经网络算法在上位机上进行视觉重建并显示。在多种测试场景和下游任务的测试结果验证了所提神经形态成像算法的有效性和系统的优越性。

综上所述，本文针对神经形态视觉传感器的视觉重建问题进行了研究，从神经形态相机的成像机理和物理建模出发，探索了生物启发的浅层脉冲神经网络、深层循环脉冲神经网络、融合脉冲和事件的视觉重建网络等算法来有效地挖掘神经形态数据中的时空信息。本文工作验证了神经形态视觉重建的技术可行性，为未来高性能神经形态系统的搭建奠定了基础。

关键词：神经形态视觉传感器，图像重建，仿视网膜采样机制，脉冲神经网络，神经形态相机

Visual Reconstruction of Neuromorphic Vision Sensors

Lin Zhu (Computer Application Technology)
Directed by Prof. Yonghong Tian

ABSTRACT

Neuromorphic camera is a brand new vision sensor that has emerged in recent years. It outputs asynchronous neuromorphic data (spikes or events) with high temporal resolution by continuously sampling the light intensity. Compared with traditional vision sensor, neuromorphic camera has the advantages of high dynamic range, high time-domain resolution and low power consumption. Thus it has great potential in the field of machine vision, such as automatic driving, UAV visual navigation, industrial detection and video surveillance, especially in high-speed motion and extreme light scenes. However, the newly-emerging asynchronous neuromorphic data can not be intuitively understood by people. Therefore, how to reconstruct continuous visual images from the sparse neuromorphic data according to the continuous sampling characteristics is a key problem in this field.

This thesis technologically formulates the neuromorphic visual reconstruction problem and establishes the evaluation benchmark. On this basis, systematic studies on continuous visual reconstruction are conducted from different perspectives: At the data level, it studies the continuous visual reconstruction from the single-mode neuromorphic data of an event camera or a spike camera, to the combinations of event and spike; At the algorithm side, by focusing on the principle of biological plausibility and computational effectiveness, this thesis starts from the hand-crafted shallow spiking neural network (SNN) to the data-driven deep spiking neural network and artificial neural network (ANN). The main contributions of this thesis can be summarized as follows:

Firstly, this thesis proposes a bio-inspired spatio-temporal spiking neural network to adaptively extract the temporal and spatial features of spike data, which can asynchronously reconstruct continuous visual images based on spike data. Firstly, according to the temporal characteristics of spike data, the motion confidence matrix is proposed to model the motion probability of spikes. Then the optimization of spatiotemporal dynamic neuron extraction is modeled as a first-order Markov random field problem and solved by graph cut. Based on the results of dynamic neuron extraction, an adaptive spike refining mechanism is further proposed

to eliminate the temporal blur. Finally, the visual reconstruction layer based on LIF neurons and synaptic connection driven by spike-timing-dependent plasticity is introduced to achieve continuous visual reconstruction. To verify the effectiveness of the proposed model, this thesis constructs the first spike-based dataset for visual reconstruction, namely PKU-Spike-Recon dataset. Experimental results show that the proposed method outperforms other approaches on both qualitative and quantitative evaluation (e.g., two-dimensional entropy and sharpness evaluation), and has high robustness on both normal and high-speed scenes.

Secondly, this thesis presents a deep recurrent spiking neural network using adaptive membrane potential neurons to extract the deep features of neuromorphic data, which achieves energy-efficient continuous visual reconstruction based on spike and event data. This thesis constructs a deep spiking neural network based on leaky integrate and fire (LIF) and membrane potential neurons for supervised learning of neuromorphic data, which achieves SNN-based large-scale image regression tasks for the first time. The features of the proposed fully SNN are transmitted by binary spikes, in this case, the deep spike features are extracted by spike skip connection and the multi-scale spike feature extraction. In the last layer, the binary spike data is converted into continuous gray values through the membrane potential (MP) of neurons, and the continuous visual reconstruction result is achieved. Based on above fully SNN framework, a hybrid SNN-ANN framework is proposed to achieve better temporal receptive field and reconstruction performance. Compared with existing ANN-based approaches, the performances of our SNNs are comparable in various scenarios, such as indoor complex lighting and outdoor automatic driving scenes, while the energy consumption of the proposed SNNs is 19.36 times and 7.75 times lower than that of their corresponding ANN structures, respectively.

Thirdly, this thesis presents a multi-modal visual reconstruction network to fuse the features of spike and event data, which can effectively achieve continuous visual reconstruction of neuromorphic cameras in complex lighting scenes. From the biologically perspective, event camera and spike camera mimic peripheral and fovea of retina, respectively. Therefore, by fusing these two kinds of neuromorphic data, we can perceive the high dynamic range motion information and fine texture information at the same time Toward this end, a novel neuromorphic data representation mechanism is proposed to adaptive extract temporal information of spike and event data. Moreover, to extract the features of spike data and event data, a multi-modal network based on a feature fusion module is designed. The complementary of the two kinds of neuromorphic data can be effectively achieved by the proposed network. To

verify the effectiveness of the proposed model, this thesis builds a neuromorphic sampling system which combines a spike camera and an event camera. A multi-modal neuromorphic dataset is constructed by the sampling system, namely the PKU-Retina-Recon dataset. Experimental results show that the proposed method can effectively combine spike and event data to reconstruct high dynamic range scenes. The SSIMs evaluation in high-speed and low illumination scenes are 21% and 16% higher than that of event-based models and spike-based models, respectively.

Fourthly, this thesis constructs a neuromorphic continuous visual reconstruction system for high-speed moving objects. The system includes the combination of FPGA and upper computer to achieve the applications of the proposed visual reconstruction models. So far, most of the traditional high-speed imaging systems are based on frame-based paradigm, thus suffer the contradiction between "mass data" and "limited bandwidth". In this thesis, we build a neuromorphic visual reconstruction system to efficiently reconstruct continuous visual information from spatiotemporal spike data, which has the advantages of high temporal resolution, less data redundancy, and low power consumption. Specifically, the system simplifies the dynamic neuron extraction and adaptive threshold mechanisms of the proposed three layer SNN, so that it can be deployed on FPGA and run in real time. Meanwhile, in order to obtain higher quality reconstruction results, two deep SNNs are designed to reconstruct and display images on the upper computer. The results on multiple challenging scenarios (e.g., high speed bullet and UAV) and downstream tasks (e.g., depth estimation and object detection) demonstrate the effectiveness of the proposed visual reconstruction models and the superiority of the constructed system.

In conclusion, this thesis focuses on the visual reconstruction of neuromorphic cameras and conducts systematic researches on both spike and event data. Starting from the sampling mechanism and physical modeling of neuromorphic cameras, this thesis explores bio-inspired shallow SNN, deep recurrent SNN, and the multi-modal fusion network based on spike and event data, to effectively process the neuromorphic data. The proposed models effectively mine the spatiotemporal features of neuromorphic data. Overall, this thesis demonstrates the technical feasibility of a neuromorphic visual reconstruction system, and lays a foundation for the construction of high-performance neuromorphic system in the future.

KEY WORDS: Neuromorphic vision sensor, image reconstruction, retina-like visual sampling, spiking neural network, neuromorphic camera