摘要

视觉系统是大脑神经系统的一个重要组成部分，人类从外界获得的信息约70%以上来自于视觉系统。视皮层及直接关联区域约占大脑皮层的三分之一，不仅负责视觉信息的提取、处理和整合，同时也是学习、记忆、决策和情感等高级脑功能形成的基础。相比于计算机视觉算法，人脑视觉系统在很多方面有独特的优势，例如人脑可以毫不费力地识别成千上万种物体，但是物体识别对于机器视觉来说却是一个富有挑战性的难题。生物视觉系统经过数亿年的进化，其中的眼睛、视觉通路和视皮层等结构、功能和工作机制都非常精妙，能够在消耗极少能量的同时快速完成许多复杂任务。生物视觉系统会在复杂的外界环境中提取出有意义的模式，而包含丰富时空信息的脉冲信号是视觉系统信息处理的基础。如何有效借鉴生物巧妙的视觉系统，实现低功耗、低延迟和高鲁棒性的类脑视觉计算是人工智能发展的重要途径之一。

为实现更具生物视觉系统特性的高效机器视觉，将围绕脉冲视觉中的编码及处理展开研究。首先探索视网膜编码视觉刺激的工作机理这个科学问题，再面向仿生物视网膜的视觉传感器研究如何构建基于脉冲的超高速视觉算法。本文的主要贡献点包括：

1. 提出了基于卷积神经网络的简单刺激单细胞编码模型，揭示了卷积神经网络与生物视网膜环路的相似性，可精确预测视网膜对白噪声和自然图片刺激的脉冲响应。无论是在不同细胞数据之间的迁移学习，还是在不同统计特性的刺激之间迁移，具有良好视网膜响应预测性能的CNN模型都表现出了较好的泛化能力。此外，在模拟单个细胞编码功能时，随着网络层数或是神经元个数的增加网络性能会趋于饱和。相比传统生物物理模型，模型拟合性能对数据样本量和噪声的影响都更加鲁棒。揭示了一个三层的卷积神经网络可相当于一个视网膜神经元。
2. 提出了基于卷积循环神经网络的复杂刺激群体细胞编码模型，揭示了网络的循环连接结构是影响视网膜编码的关键因素，可高精度地预测大规模视网膜神经节细胞对动态自然场景的响应，并同时学习出各个神经节细胞感受野的形状与位置。所提出的卷积循环编码网络除了在结构上更加接近视网膜，还可以使用更少的参数学习出精度更高的编码模型。这一模型不仅具有生物学价值，而且对设计新一代脉冲视觉模型、芯片乃至研制视网膜假体都具有重要意义。
3. 提出了基于短时程可塑性模型的脉冲视觉图像重构算法，仿照大脑中的短时程可塑性机制动态记录脉冲流的时间特性，在不受统计时间窗口长度的限制下，

既保留脉冲相机低时间延迟的特性，又可获得低噪声、少动态模糊的高质量重构图像。所提出方法可根据脉冲相机产生的脉冲流特性，动态调整神经元的状态，从而推导出场景像素值。此外，还分析了如何权衡地减少静态区域的噪声和动态区域的模糊，及如何减少脉冲相机读出信号时产生的噪声。实验表明，所提出的重构算法在主观和定量的评价结果上都优于现有的脉冲相机重构算法。

4. 提出了超高速脉冲视觉检测跟踪框架，能够以无监督地方式直接处理脉冲流信息，在线动态调节脉冲神经网络状态以实现对多个超高速目标的连续跟踪。通过依次引入动态适应层、运动估计模块和基于脉冲神经元的物体检测跟踪层，构建了三个面向脉冲视觉的多目标检测跟踪算法。其中第一种算法的复杂度较低，主要面向类似于监控或者定点监测等相机固定的高速场景。而后两种算法中包含基于脉冲神经元的运动估计模块，可以适用于包含相机自运动的场景。这些基于脉冲信号的检测跟踪算法可便于移植至神经形态计算平台，同时发挥神经形态视觉传感器与计算平台的低功耗、低延迟优势。

综合对生物视觉系统工作机理的探索，与面对脉冲视觉的超高速视觉算法的研发，设计了一套超高速脉冲视觉演示验证系统。从系统实现的角度对上述面向脉冲视觉的创新性方法及其特点和适用范围进行了整理和归纳，以便用户可以根据具体的场景需求获取超高速场景中的关键信息，如进行超高速运动/变化过程的科学观测，超高速运动目标的追踪等。最后，以超高速目标打击和高速激光追踪作为本工作所提出系统的典型应用案例，介绍如何实现从本文方法到实际超高速应用的技术转化。

关键词：生物视觉系统，视网膜编码模型，脉冲相机，超高速脉冲视觉算法，脉冲神经网络

Biological Encoding Models and Ultra-High-Speed Algorithms for Spike Vision

Yajing Zheng (Computer Science)
Directed by Prof. Tiejun Huang

ABSTRACT

The visual system is an important part of the brain's nervous system. About 70% of the information that humans obtain from the outside world comes from vision. The human visual cortex and its directly related areas account for about one-third of the cerebral cortex. It is not only responsible for the extraction, processing and integration of visual information, but also the basis for the formation of advanced brain functions such as learning, memory, decision-making, and emotion. Compared with computer vision algorithms, the human brain visual system has unique advantages in many aspects. For example, for the task of object recognition, the human brain can effortlessly identify thousands of objects, but it is a challenging problem for machines. The biological visual system has evolved over hundreds of millions of years. Its functions such as eyes, visual pathways and visual cortex, as well as its system' s working mechanism, are very sophisticated. The perception ability of even small insects in some scenes is much better than most of the machine vision system in the world. How to effectively learn from the design of biological ingenious vision system and realize a more advanced machine vision paradigm is still a problem to be studied.

In order to realize advanced machine vision with more characteristics of the biological vision system, this paper first explores the scientific question of "how does the biological vision system see the world", and then studies "how to help machines see the world" for the retina-inspired neuromorphic vision sensors. The main contributions of this paper include:

1. Proposing a single-cell encoding model for simple stimulus based on a convolutional neural network, revealing the similarity of convolutional neural network with biological retinal circuit, and can accurately predict retinal spiking responses to white noise and natural picture stimuli. Experimental results show that the CNN model with good performance of retinal response prediction exhibits better generalization ability, whether in transfer learning between different cells or stimuli. In addition, when

simulating the encoding function of a single cell, the network performance will tend to be saturated with the increase in the number of network layers or neurons, revealing that a three-layer convolutional neural network can be equivalent to a retinal neuron - the smallest neural network in the biological visual system.

2. Proposing a population cell encoding model for complex stimulus based on convolutional recurrent neural network, which can predict the response of large-scale retinal ganglion cells to dynamic natural scenes with high accuracy, and can learn the receptive field of each ganglion cell, unraveling the coding principle of visual neruons for dynamic scenes and provide a way of using recurrence for understanding isual computing. In addition to being more similar to the retina in structure, the proposed convolutional recurrent coding network can also learn a higher-accuracy coding model with fewer parameters. This model is not only of biological value, but also has important significance for the design of a new generation of spiking vision models and chips, and even the development of retinal prostheses.
3. Proposing a high-speed image reconstruction algorithm for spiking cameras based on the short-term plasticity model, which imitates the short-term plasticity mechanism in the brain to dynamically record the temporal characteristics of the spike stream, and retains the low time delay of the spiking camera without being limited by the length of the time window. It can obtain high-quality reconstructed images with low noise and less motion blur. In view of the characteristics of the spike streams generated by spiking camera, this paper will analyze how to deduce the spatial pixels according to the state of the spiking neuron without being affected by the parameter setting of the statistical time window. In addition, this paper analyzes the trade-off between reducing noise in static regions and blurring in dynamic regions, and how to reduce the noise generated when the spiking camera reads out the signal. Experiments show that these improvements are of great benefit to the reconstruction of ultra-high-speed scenes. Experiments show that the proposed algorithms outperform the existing spiking camera reconstruction methods in both subjective and quantitative evaluation results.
4. Proposing an ultra-high-speed spiking visual detection and tracking framework, which can directly process the spike flow information in an unsupervised manner, and dynamically adjust the state of the spiking neural network online to achieve continuous tracking of multiple targets. By introducing the dynamic adaptation layer, the motion estimation module, and the object detection and tracking layer based on spiking neurons

in turn, this paper constructs three multi-target detection and tracking algorithms for spiking vision, namely Spike SORT, Spike SORT2, and SNNTracker. Among them, the algorithm complexity of Spike SORT is low, and it is mainly oriented to high-speed scenarios with fixed cameras such as surveillance or fixed-point surveillance. The Spike SORT2 and SNNTracker include spike-based motion estimation modules, which can also be applied to scenes containing the camera's ego-motion. The proposed algorithms can be easily ported to neuromorphic computing platforms, while taking advantage of the low power consumption and low latency of neuromorphic vision sensors and computing platforms.

Based on the exploration of the working mechanism of the biological vision system, and the research and development of the ultra-high-speed vision algorithm for spike vision, a set of ultra-high-speed spiking vision demonstration and verification system is designed in this paper. From the perspective of system implementation, the above-mentioned innovative methods for spike vision, their characteristics and application scope are sorted and summarized, so that users can obtain key information in ultra-high-speed scenes according to specific scene requirements, such as ultra-high-speed motion/change scientific observation of the process, tracking of ultra-high-speed moving targets, etc. Finally, this paper takes ultra-high-speed target strike and high-speed laser tracking as typical application cases of the system proposed in this work, and introduces how to realize the technical transformation from the method in this paper to practical ultra-high-speed applications.

KEYWORDS: Biological Visual System, Retinal encoding model, Spiking camera, Ultra-high-speed spiking vision, Spiking neural networks