超高速脉冲相机的成像方法综述

赵睿　熊瑞勤　董彦辰　赵菁　郑雅菁　黄铁军

(北京大学 计算机学院 多媒体信息处理全国重点实验室 北京 100871)

摘　要　高速动态场景的清晰成像是工业检测、自动驾驶等视觉应用的关键需求。传统帧式相机受限于全局曝光机制，在高速场景下面临短曝光引入噪声与长曝光导致运动模糊的两难困境，且固定帧率难以捕捉超高速瞬态过程。神经形态视觉传感器，特别是模拟视网膜中央凹工作机制的脉冲相机，通过各像元独立、异步的“积分－发放”机制连续输出脉冲流记录光学场景，为突破传统成像的瓶颈提供了新思路。本文系统综述了超高速脉冲相机的成像方法，首先简要介绍了脉冲相机的工作原理与脉冲流特性；然后梳理了基于单个脉冲相机的成像算法：基础灰度影像重建方法涵盖基于时空滤波、生物神经元机制模拟及端到端深度学习三大技术路线，旨在有效利用脉冲流的时空上下文信息，兼顾噪声抑制与运动处理；随后，本文进一步拓展至超分辨率重建、彩色影像重建、高动态范围成像、特殊场景成像及结合神经辐射场或三维高斯泼溅的三维成像等细分方向。其次，本文探讨了融合脉冲相机与传统帧式相机或事件相机的多传感器成像算法，重点阐述了面向高帧率、高动态成像与模糊消除的融合方案，实现了不同传感器间的优势互补。最后，总结了当前研究进展，并对未来的相关研究进行了展望。

关键词　脉冲相机；成像算法；时空信息融合；多传感器融合

中图法分类号　TP391

Review of imaging methods for ultra-high-speed spike camera

ZHAO Rui　XIONG Rui-Qin　DONG Yan-Chen　ZHAO Jing　ZHENG Ya-Jing　HUANG Tie-Jun

(State Key Laboratory of Multimedia Information Processing, School of Computer Science, Peking University, Beijing 100871)

Abstract　High-fidelity imaging of high-speed dynamic scenes is a critical imperative in various advanced visual applications, ranging from industrial inspection to autonomous driving systems. However, traditional frame-based cameras are fundamentally constrained by their global exposure mechanisms and fixed frame rates, inevitably facing a dilemma in high-speed scenarios: short exposure times introduce significant noise due to insufficient photon accumulation, while long exposures result in severe motion blur, and the discrete nature of frame capture leads to "blind spots" between frames, causing the loss of crucial transient information. To break through these bottlenecks, neuromorphic vision sensors have emerged as a revolutionary solution, particularly the spike camera, which mimics the working mechanism of the retinal fovea. Unlike event cameras that only record intensity changes, the spike camera records optical fields continuously through an asynchronous, pixel-independent "integrate-and-fire" mechanism, where each pixel accumulates photons into voltage and fires a binary spike upon reaching a threshold. This paradigm allows for the output of continuous spike streams with ultra-high temporal resolution, offering a novel approach to capturing rapid optical changes without the limitations of a fixed clock. This paper presents a systematic and comprehensive review of imaging methods for ultra-high-speed spike cameras, aiming to bridge the gap between raw binary spike streams and high-quality

———————————————

赵睿，博士，主要研究领域为计算摄像、计算机视觉. 熊瑞勤，博士，研究员，计算机学会(CCF)会员，主要研究领域为图像视频处理、深度学习、计算成像. 董彦辰，博士研究生，主要研究方向为计算机视觉、计算摄影学和图像复原.　赵菁，博士，主要研究领域为计算摄像、计算机视觉.郑雅菁，博士，计算机学会(CCF)会员，主要研究领域为类脑计算、神经形态视觉.黄铁军（**通信作者**），博士，教授，计算机学会(CCF)会士，主要研究领域为视觉信息处理、类脑智能等.

visual reconstruction. The review begins by elucidating the fundamental working principles of the spike camera and the unique characteristics of its output data, highlighting its ability to preserve both high-speed motion flux and static texture information effectively. Subsequently, the paper systematically categorizes and details imaging algorithms based on single spike cameras. For basic grayscale image reconstruction, three primary technical routes are analyzed in depth: methods based on spatial-temporal filtering which utilize motion alignment to aggregate context and mitigate stochastic noise; bio-inspired methods that simulate neural mechanisms like Spike-Timing-Dependent Plasticity (STDP) and Short-Term Plasticity (STP) to adaptively process temporal correlations; and end-to-end deep learning approaches designed to learn complex non-linear mappings for noise suppression and detail recovery, including recent self-supervised frameworks. Building upon basic reconstruction, this review further explores advanced imaging tasks that extend the capabilities of spike cameras. These include super-resolution reconstruction which converts high temporal resolution into spatial detail, color image reconstruction which addresses the intricate challenges of demosaicing asynchronous binary streams from Bayer-pattern sensors, and High Dynamic Range (HDR) imaging which leverages rolling-mixed-bit mechanisms to resolve the trade-off between highlight saturation and low-light noise. Notably, the review also covers cutting-edge developments in three-dimensional imaging, discussing how continuous spike streams are integrated with Neural Radiance Fields (NeRF) or 3D Gaussian Splatting to achieve high-quality 3D scene representation from fast-moving viewpoints. Additionally, the paper examines applications in special scientific scenarios, such as recording ultra-fast phase transitions and microfluidic droplet analysis. Furthermore, recognizing the spatial resolution limitations of current spike camera hardware, this paper critically discusses multi-sensor fusion algorithms that combine spike cameras with traditional frame-based cameras or event cameras. These fusion strategies are specifically highlighted for their applications in high-frame-rate video interpolation, dynamic range expansion, and motion deblurring, demonstrating how hybrid systems can achieve "high temporal resolution, high spatial resolution, and high dynamic range" simultaneously by leveraging cross-modal complementarity. Finally, the paper summarizes the current state of research, identifies existing challenges such as the lack of large-scale real-world datasets, and proposes future research directions, including semantic-driven imaging, spike-based light field imaging, and hardware-algorithm co-design, to pave the way for the practical deployment of spike vision technology.

Key words spike camera; imaging method; spatial-temporal information fusion; multiple-sensor fusion

1 引言

在工业检测、自动驾驶等计算机视觉应用中，实时、精确地捕获高速视觉图像并处理其蕴含的视觉信息，是视觉系统发展的重要方向之一[1]。当前广泛应用的基于帧采样原理的视觉传感器因其固有的全局同步曝光机制，在高速场景中陷入一种两难困境：短曝光时间虽能减少高速运动导致的动态模糊，却因光子累积不足引入显著噪声，尤其在低光环境下成像信噪比急剧下降；若延长曝光时间，虽能通过充分累积光子降低噪声，却会因场景中物体的位移产生运动模糊，导致细节丢失[2]。这种以固定时间间隔抓取静态影像的方式，本质上牺牲了连续时间维度上的信息完整性，致使瞬变的光强过程与超高速动作细节难以被有效记录。此外，传统相机的固定帧率（通常为 30-60Hz）使其无法捕捉帧间间隔内发生的高速瞬态过程，时域信息的丢失限制了视觉系统对动态场景的完整理解。在诸如自动驾驶中车辆紧急避障与遭遇强烈逆光等高速或光照条件剧变的实际复杂场景下，传统帧采样式相机成像能力的瓶颈被充分暴露，其难以捕获清晰的影像，严重影响了视觉感知系统的精度与可靠性。

为突破传统成像范式的瓶颈，神经形态视觉传感器[3-5]应运而生，其仿照生物视网膜机理，通过各像元异步工作实现对光学场景的连续记录。目前神经形态相机主要可被划分为脉冲相机（Spike Camera）[1,6]和事件相机（Event Camera）[7-8]两种类型。其中事件相机模拟了视网膜的外周（Periphery），其连续地对场景光照的变化进行记录，当光照强度

的变化在对数域超过阈值时，对应的像元就发放一个包含坐标、时间戳与光强变化方向的“事件”。而脉冲相机模拟的是视网膜的中央凹（Fovea），其每个像元独立地对光子进行连续的累积，当累积值达到阈值时发放一个脉冲，并立即开始下一个脉冲的累积。相较于事件相机对场景光照变化的记录，脉冲相机基于其“累积-发放”的工作机制，使场景中静态和动态区域的信息都能被有效保留，为解决传统相机的固有缺陷提供了全新思路。

凭借脉冲相机的上述优势，研究者们开展了一系列研究工作，围绕该相机的视觉任务研究日益丰富[9]，涵盖光流估计[10-13]、目标检测[14-15]、深度估计[16-17]等多个方向。然而，这些视觉任务的实现，均依赖于从脉冲流中解析出光强信息。脉冲式成像旨在从脉冲相机输出的脉冲流中解析得到清晰的影像，其作为连接原始脉冲信号与高层视觉理解的核心桥梁，是脉冲视觉技术落地的基础与前提。

从算法层面看，脉冲相机的成像过程是对脉冲流中蕴含的光强信息进行解析的过程。早期研究主要通过统计脉冲流中的时序特征，例如 TFI 和 TFP 方法[18]分别利用脉冲间隔与光强的反比关系与一个时间窗口内的脉冲数量对各像素的值进行推断，这类方法只考虑了单个像素，没有利用连续脉冲流中丰富的时间和空间上下文信息，难以平衡噪声抑制与运动模糊的矛盾。后续的方法对连续脉冲流中的时空上下文利用方案展开了探索。例如基于运动估计将不同时刻的脉冲信息进行对齐[19]、基于卷积神经网络进行时空特征提取[20]、基于脉冲神经网络（Spiking Neural Network, SNN）[21-23]模拟生物神经元的脉冲传递机制进行时空信息聚合[24]。对于这些方法，本文将其归类为基于滤波的方法、基于生物神经元的方法与基于深度学习的方法三类。

近年来，脉冲相机成像的相关研究工作在不同层面上进行了拓展，从等分辨率重建拓展至超分辨率重建[25]、从灰度相机拓展至彩色相机[26]、从一般化的场景拓展至特殊场景[27]等细分方向。研究者们还进一步探索了脉冲相机在三维场景感知中的潜力，通过将连续时间维度的脉冲信息转化为空间维度的思路，实现了基于神经辐射场（Neural Radiance Field）[28]或高斯泼溅（Gaussian Splatting）[29]等范式的三维成像。此外，除了单相机成像之外，脉冲相机与传统或事件相机的融合算法成为了重要的研究方向[30]，通过互补两者在时间分辨率与空间分辨率上的优势，有效解决了传统相机的运动模糊、过曝、欠曝等问题，提升了复杂场景下的成像质量。

目前，与本文相关的综述工作主要从不同视角展开：如文献[31]主要聚焦于脉冲连续摄影的物理机制与系统验证；文献[4]更多面向事件相机等仿视网膜传感器，脉冲相机仅作部分介绍；文献[2]侧重于脉冲视觉的原理、硬件实现及各类下游视觉任务；文献[32]主要分析了脉冲相机在自动驾驶中的应用前景。与上述工作不同，本文以“如何从脉冲相机输出的脉冲流获得高质量影像”为核心主线，系统梳理了各类脉冲成像算法。

如图 1 所示，本文围绕脉冲相机的成像算法展开综述，整体结构按照“从单一脉冲相机到多传感器系统、从基础重建到任务扩展”的技术脉络进行组织。首先对脉冲相机及其工作原理进行简介；进而从成像视角分析脉冲相机成像的关键问题与挑战；然后按成像任务类型对脉冲相机的成像算法进行梳理，涵盖基础影像重建、超分辨率重建、彩色影像重建、HDR 成像和三维影像重建等关键技术；继而讨论融合脉冲相机的多传感器成像系统，梳理传感器融合的技术路线与典型方案；最后总结当前研究面临的挑战并展望未来发展方向，以期为脉冲相机成像算法的研究与应用提供参考。

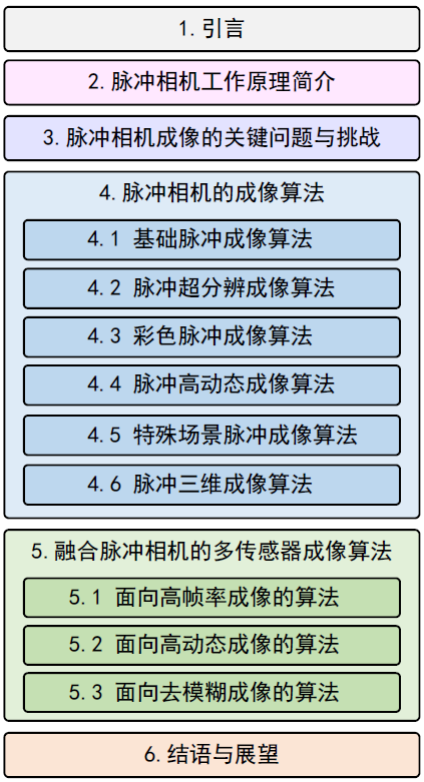

图 1　本文的组织架构

2 脉冲相机的工作原理简介

2.1 脉冲的产生过程

脉冲相机单个像元的构造及运作机制如图 2 所示。每个像元主要由光子接收器、积分器和比较器构成。来自场景的光子被接收后，转化为积分器中的电势。当积分器里的电势累积到设定的发放阈值时，就会立即发放一个脉冲，同时将累积的电势清零。该过程亦可从光子累加的角度理解：即当到达的光子数量达到发放阈值时，便会产生脉冲并重置积分器。需要注意的是，在脉冲相机的工作过程中，各个像元的光子累积和脉冲发放都是独立进行的，也即每个像元的 “积分 - 发放” 过程是异步的。基于上述机制，可建立数学模型：设空间坐标$\mathbf{x}$处的像元在时刻t时，单位时间内到达的光子数量期望为$\mathbf{L}(\mathbf{x},t)$，积分器中的累积值可以被表示为[19]：

$$\mathbf{A}(\mathbf{x},t)=\int_0^t \alpha\mathbf{L}(\mathbf{x},t)\ \mathrm{d}\tau\ \ \mathrm{mod}\theta, \tag{1}$$

其中$\mathbf{A}$为积分器中的累积值，α为光子转化效率，也即电路将实际到达的光子数量转化为积分器中累积值的效率，θ为脉冲发放阈值。在真实世界的摄像中，光子到达遵循泊松分布，考虑这一效应后，上式可被推广为：

$$\mathbf{A}(\mathbf{x},t)=\int_0^t \alpha\mathcal{P}(\mathbf{L}(\mathbf{x},t))\ \mathrm{d}\tau\ \ \mathrm{mod}\theta, \tag{2}$$

其中$\mathcal{P}$代表泊松采样。

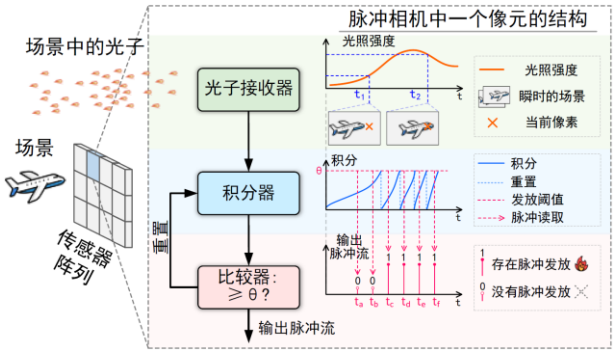

图 2 脉冲相机的像元结构与工作原理[33]

2.2 脉冲的读取过程

上述内容阐述了累积光子与发放脉冲的过程。在脉冲相机工作过程中，各像元的“积分 - 发放”过程独立进行，而脉冲的读出受时钟信号控制。设时钟信号的周期为T_r，则脉冲相机每隔时间T_r将所有像元的脉冲读出。若在当前读出时刻与上一读出时刻之间，存在脉冲发放，读出为 1，否则读出为 0。设脉冲流为$\mathbf{S}(\mathbf{x},k)$，其中k表示脉冲读取的次数，则该信号的值可以被表示为：

$$\mathbf{S}(\mathbf{x},k)=\begin{cases}1, \text{ if } \exists\ t\in((k-1)T_r,\ kT_r] \text{ s.t. } \mathbf{A}(\mathbf{x},t)=\theta,\\ 0, \text{ if } \forall\ t\in((k-1)T_r,\ kT_r] \text{ s.t. } \mathbf{A}(\mathbf{x},t)<\theta.\end{cases} \tag{3}$$

可见，对于包含M个读出时刻的脉冲流而言，其数据型式为$H\times W\times M$的二值矩阵。对于基于脉冲相机的成像而言，其关键即为从上述二值脉冲流中解析得到指定时刻的高质量影像。

3 脉冲相机成像的关键问题与挑战

脉冲相机在走向实际应用过程中，需要解决源于其独特工作原理的一系列问题，这些挑战引领了成像算法的发展。

3.1 脉冲流的固有特性与挑战

脉冲相机从连续光强信号中编码离散二值脉冲流。即使光照稳定，脉冲间隔仍会因泊松光子到达、热噪声和量化效应而波动。从二值脉冲流中解码光强面临两点难题：其一，二值脉冲仅代表“阈值触发”，重建为典型欠定逆问题；其二，像元独立异步发放，虽避免运动模糊，却使跨像素与跨时间的对齐与融合困难，尤其在动态场景中增加了运动估计的不确定性。

3.2 脉冲相机基础成像的关键问题

在脉冲成像的早期研究中，核心问题是如何在噪声抑制与运动模糊之间取得平衡。窗口脉冲计数法（TFP）通过时间窗口统计脉冲数量来估计光强，能有效抑制随机噪声，但在高速运动场景下易产生严重模糊；脉冲间隔法（TFI）则利用相邻脉冲间隔推断光强，可减少模糊，但对噪声非常敏感，稳定性不足。由此可见，仅靠时域聚合无法兼顾噪声与清晰度，关键挑战在于超越单像素的孤立处理，充分利用脉冲流的时空上下文，在运动对齐的同时实现有效聚合。这一需求推动了后续滤波、神经元模拟和深度学习等方法的发展。

3.3 拓展应用面临的关键挑战

在克服基础成像的挑战后，脉冲相机成像技术开始向更广泛的应用领域延伸，每个领域都面临着新的、独特的挑战。

超分辨率成像：核心挑战在于将脉冲流的高时间分辨率转化为空间分辨率，不仅依赖运动带来的亚像素信息聚合，还要解决从稀疏时序采样中恢复高精度空间细节的逆问题。

彩色成像：传统的彩色相机通过彩色滤光阵列（Color Filter Array, CFA）实现单像素对单一颜色通道的采样，脉冲相机的彩色成像同样遵循此原理。脉冲彩色成像核心挑战在于高速动态场景下的去马赛克与颜色重建，需要跨通道时空对齐与协同去噪，避免色彩伪影和混叠。

高动态范围成像：由于积分－发放机制，高光区易产生脉冲累积过饱和的问题，也即一个读取周期内的累积值超过阈值，从而丢失信息，低光区则脉冲稀疏噪声大，难以在单一传感器下兼顾明暗细节。其核心挑战在于，如何在单一传感器架构下同时兼顾高光与低光区域的精细细节。

三维成像：将脉冲流用于三维场景重建，其挑战在于如何将二维脉冲流整合为一致的三维表示，并结合神经辐射场或 3D 高斯泼溅等范式，从稀疏动态数据中学习几何与光度信息。

3.4 多传感器融合的关键问题

由于数据带宽的限制，单一脉冲相机通常空间分辨率相较于传统相机较低，难以满足精细静态纹理需求；传统帧式相机虽具高空间分辨率，但动态范围有限；事件相机具高动态范围，却无法提供静态区域完整纹理。这些优势互补的特性推动了多传感器融合的发展。其核心挑战在于数据对齐与最优信息融合，不同传感器在时间戳、空间分辨率和物理形态上差异显著。准确地对齐异步脉冲流与固定帧率的 RGB 图像和事件流，尤其在未严格预校准的光学设置下，是一项复杂任务。此外，设计智能融合算法，避免伪影并实现高帧率、高动态范围及去模糊等高级成像任务，是该领域的关键问题。

4 脉冲相机的成像方法

脉冲相机基于上述“积分－发放”机制输出的脉冲流虽蕴含丰富的时空信息，却以二值化、时序化的特殊形式存在，无法直接被人眼或传统视觉算法解读。因此，从脉冲流中解码出光照强度，是实现高质量影像重建、体现脉冲相机成像优势的关键环节。在脉冲相机的成像研究中，单相机系统作为最基础的形态，其算法发展直接体现了脉冲成像技术的核心突破。本节按“先基础、后扩展”的递进关系组织：4.1 节讨论单相机的基础灰度重建，作为后续方法的共性基础；4.2－4.5 节在此基础上，面向不同任务需求分别扩展超分辨率、彩色成像、高动态范围成像和特殊场景成像，彼此在应用方向上并列、在技术上共享基础重建框架；4.6 节则进一步结合神经辐射场和三维高斯泼溅等模型，将前述二维成像扩展到三维场景表示。本节重在梳理不同技术路线如何应对脉冲流解码中的噪声抑制、运动对齐与时空关联建模等关键挑战，回顾脉冲相机成像能力的演进与突破。

4.1 基础脉冲成像算法

在本文中，基础脉冲成像算法指从脉冲流中重建出与输入脉冲流空间尺寸一致的灰度二维场景影像的算法，该类任务不涉及高动态范围扩展、多视点影像渲染、多维度信息融合等进阶需求。其聚焦于解决脉冲流解码中最本质的光照强度恢复问题，是脉冲成像技术的基础与核心。

与仅对光照变化敏感的事件相机成像方法相比，各类脉冲成像算法普遍利用了脉冲流同时保留静态纹理与动态运动的信息：一方面，在静态区域通过长时积分抑制噪声、提升细节保真度；另一方面，在动态区域结合脉冲高时间分辨率进行运动对齐与模糊抑制。这种“静态+动态”统一建模能力，能够充分利用脉冲相机对场景记录完备的优势。

早期针对脉冲相机成像的探索主要基于简单的统计模型展开，典型的两大方法为 TFP（Texture from Playback）和 TFI（Texture from Interval），本文将这两种方法分别命名为窗口脉冲计数法和脉冲间隔法。如图 3 所示，这两种方法关注于单个像素沿时间轴的统计信息。其中 TFP 基于对一定时间窗口内各像素的脉冲发放频率进行光强推断，如果设拟重建的时刻为t_{key}，时间窗口的半径为r_{win}，则 TFP 可以被表示为：

$$\mathbf{I}(\mathbf{x},t_{\mathrm{key}})=\frac{1}{2k+1}\sum_{k=t_{\mathrm{key}}-r_{\mathrm{win}}}^{t_{\mathrm{key}}+r_{\mathrm{win}}}\mathbf{S}(\mathbf{x},t), \tag{4}$$

其中r_{win}为算法的超参数，针对不同的场景，需要设置不同的r_{win}来获得较优的结果。TFI 基于拟推断的像素在拟重建时刻所处于的脉冲积分周期进行光强推断，该方法将光强可化为脉冲积分周期的反比，设$\mathbf{P}(\mathbf{x}_0,t_{\mathrm{key}})$和$\mathbf{N}(\mathbf{x}_0,t_{\mathrm{key}})$分别在是$t_{\mathrm{key}}$时刻，空间位置$\mathbf{x}_0$处的前一个脉冲和后一个脉冲的发放时间，这两者与脉冲流$\mathbf{S}$的关系可以被表示为[19]：

$$\mathbf{P}(\mathbf{x},n)=\max\{k\mid\mathbf{S}(\mathbf{x},k)=1,\ k<n\}, \tag{5}$$

$$\mathbf{N}(\mathbf{x},n)=\min\{k\mid\mathbf{S}(\mathbf{x},k)=1,\ k\geq n\}. \tag{6}$$

在此基础之上，便可基于脉冲的积分周期，也即发放间隔，对光强进行推断：

$$\mathbf{I}(\mathbf{x},t_{\mathrm{key}})=\frac{1}{\mathbf{N}(\mathbf{x},n)-\mathbf{P}(\mathbf{x},n)} \tag{7}$$

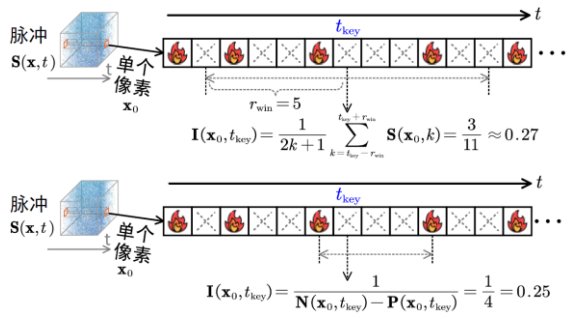

图 3 窗口脉冲计数法（TFP）与脉冲间隔法（TFI）脉冲成像算法示意图

图 4 展示了基于上述两种方法得到的脉冲成像结果，其中“列车行驶”为在一辆高铁上拍摄另一辆对向行驶的高铁所得，相对运动速度约为 700 公里每小时；“风扇旋转”的拍摄内容为每分钟 2600 转的电风扇。对于图 4 中的结果而言，TFP 的窗口半径为 12。可见，TFP 和 TFI 这两种方法虽能直观地将各像素沿时间轴的脉冲统计规律与光强建立关联，但其成像效果较为有限。TFP 在高速运动场景中易产生模糊；TFI 只考虑一个脉冲的间隔，容易产生较大的像素值波动。

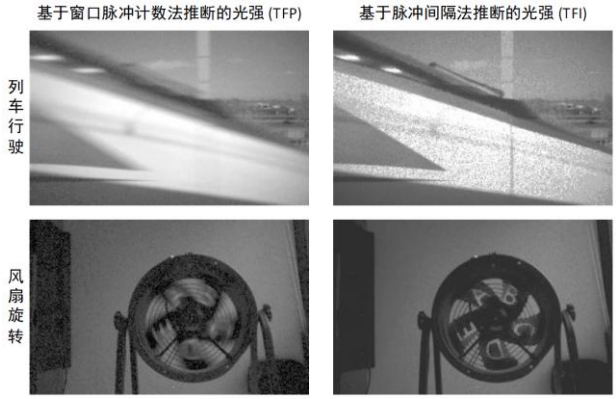

图 4 基于窗口脉冲计数法（TFP）与脉冲间隔法（TFI）得到的脉冲成像结果

考虑到 TFP 和 TFI 对脉冲流中时空上下文利用的有限性，后续研究者们从不同角度出发，探索了更有效的脉冲流时空上下文利用方式，以突破 TFP 和 TFI 的局限性，主要可归纳为以下三类技术路线：（a）基于滤波的时空信息融合方法、（b）基于生物神经元机制的时空建模方法、（c）基于端到端深度学习的时空信息融合方法，三类方法的对比总结如表 1 所示，以下分别展开介绍。

（a）基于滤波的时空信息融合方法

基于滤波进行时空信息融合的脉冲基本成像方法从不同的角度体现了对时空信息进行利用的思路。在脉冲相机连续记录场景的过程中，场景中的物体运动会导致同一物体在不同时刻成像于不同像素上，换言之，对于同一像素而言，在时间不同时，其对应的场景内容是不同的。连续脉冲流蕴含丰富的时序信息。为利用这些信息提升成像质量，需针对拟重建时刻的每个像素，估计其在邻近时刻对应场景点的位置，从而聚合同一物体在不同时刻的采样信息。该过程本质上是进行运动估计。

对于二维影像而言，像素级的运动估计一般通过光流[34]进行刻画。光流也即影像不同帧之间的像素级对应关系，如果设光流 $\mathbf{w}(\mathbf{x})$ 表示以图像 $\mathbf{I}_1$ 为起点、图像 $\mathbf{I}_2$ 为终点的运动，则该关系可被表示为：

$$\mathbf{I}_2(\mathbf{x}+\mathbf{w}(\mathbf{x}))\rightarrow\mathbf{I}_1(\mathbf{x}), \tag{8}$$

其中 $\rightarrow$ 代表像素级对应。基于上述光流，便可知图像 $\mathbf{I}_1$ 中的各个像素在图像 $\mathbf{I}_2$ 中的位置。

如图 5 所示，MAF[35]首先提出基于光流估计，找到拟重建时刻与周围邻近时刻不同帧的运动对齐关系，进而沿着运动方向进行基于高斯滤波的信息聚合，从而实现时间上下文的自适应利用。该方法从脉冲的间隔域出发，可以被公式化地表达为：

$$\mathbf{I}(\mathbf{x},t_{\text{key}})=\sum_{k=t_{\text{key}}-r_{\text{win}}}^{t_{\text{key}}+r_{\text{win}}}\frac{g_{t_{\text{key}}}^{k}\cdot\mathbf{I}_{\text{TFI}}(\mathbf{x}+\mathbf{w}(\mathbf{x},k|t_{\text{key}}),k)}{2k+1}, \tag{9}$$

其中 $\mathbf{w}(\mathbf{x},k|t_{\text{key}})$ 为以时刻 t_{key} 为起点、时刻 k 为终点的光流，可以通过 TFI 方法获得的影像经 Brox 光流法[36]得到；$\mathbf{I}_{\text{TFI}}(\mathbf{x},k)$ 为时刻 k 的经脉冲间隔法得到的初步估计影像；$g_{t_{\text{key}}}^{k}$ 为重建时刻 t_{key} 时，时刻 k 的影像的高斯权重：

$$g_{t_{\text{key}}}^{k}=\frac{1}{\sqrt{2\pi}\,\sigma}\exp\left(\frac{-(t_{\text{key}}-k)^2}{2\sigma^2}\right), \tag{10}$$

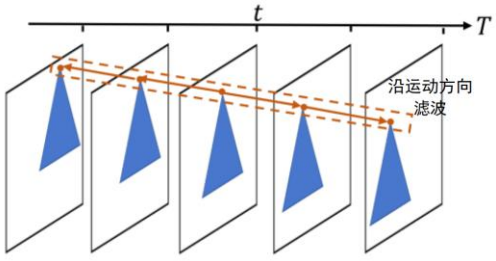

图 5 MAF[35]中提出的沿运动方向进行滤波

Zhang 等人[37]提出的 LAP 从另一种角度利用时空信息，通过度量不同时间位置上脉冲发放间隔长度的相似性构造时域滤波器权重，进而通过自适应的时域滤波聚合时空信息。设位置 $\mathbf{x}$ 在时刻 t 所处的脉冲积分周期为 $\mathbf{D}(\mathbf{x},t)$，则滤波的权重可以被表示为（省略空间坐标）：

$$\omega_{\text{LAP}}(t_0|t_{\text{key}})=\frac{\exp\left(\frac{-|\mathbf{D}(t_{\text{key}})-\mathbf{D}(t_0)|}{h\mathbf{D}(t_{\text{key}})}\right)^{\alpha}}{2\mathbf{D}(t_{\text{key}})}, \tag{11}$$

表 1 基础脉冲成像算法三大类方法的比较

方法类别	计算复杂度	噪声鲁棒性	运动敏感度	重建质量
滤波类方法	中低	依赖滤波器设计，可抑制部分随机噪声，但对脉冲丢失或错误发放较敏感	初级方法主要适用于静态或缓慢运动场景，复杂方法依赖于传统光流估计方法的精度	可获得较平滑的重建，但细节和纹理保留能力有限
生物神经元类方法	中低	通过模拟神经元工作机制可一定程度抑制噪声，但整体效果依赖具体神经元模型设计与参数标定	可通过时间常数和阈值调节等策略适配不同速度范围，但在较复杂的运动场景下表现有限	在动态区域重建效果通常优于初等滤波方法，但整体质量与模型逼真度和参数选择密切相关
深度学习类方法	高	在充分数据驱动下可对多种噪声模式进行建模，对常见噪声具有较强鲁棒性	可学习复杂运动模式与遮挡关系，对快速和非刚体运动更敏感	在空间细节恢复和时间一致性方面通常优于其他类方法

其中h与α为控制滤波权重的超参数。基于上述由相似性刻画的权重，即可对一定时间内的脉冲发放间隔进行滤波，该过程可以被表示为：

$$\bar{\mathbf{D}}(t_{\text{key}})=\frac{\sum_{k=t_{\text{key}}-r_{\text{win}}}^{t_{\text{key}}+r_{\text{win}}}\omega_{\text{LAP}}(k|t_{\text{key}})\cdot\mathbf{D}(k)}{\sum_{k=t_{\text{key}}-r_{\text{win}}}^{t_{\text{key}}+r_{\text{win}}}\omega_{\text{LAP}}(k|t_{\text{key}})},\tag{12}$$

在此基础上，可以通过光强与脉冲积分周期的反比关系得到所估计的光照强度。Wang 等人[38]在 LAP 的基础上进一步考虑了脉冲间隔的空间相似性，该研究将相似性滤波的范围从局部时域扩展到了局部时空域，也即在三维时空立方体中对脉冲间隔进行相似性滤波，进一步提升了所得影像的精度。

早期脉冲滤波方法主要基于时间窗口内的脉冲发放频率或脉冲间隔估计光照强度：前者对随机噪声更鲁棒但在运动场景中易产生模糊，后者能更准确反映瞬时光强，却对噪声更为敏感。为综合两者优势，Dong 等人[39]提出 BALI 算法，通过亮度掩码将场景划分为明暗区域，对亮区域采用脉冲窗口统计法、对暗区域采用间隔法估计光强，并结合光流估计沿运动方向对初步结果进行滤波，最终以亮度自适应的非局部空域滤波抑制暗区噪声，从而提升影像重建的质量。

上述 BALI 算法体现了多阶段重建的思想：首先基于简单的方法得到初步的影像重建结果，再通过后续的精细化处理进一步提升重建影像的质量。MAHTF[19]在沿运动方向进行滤波的基础上，进一步提出基于长短时长分层滤波的脉冲成像算法。该方法首先采用沿运动方向的滤波技术，对由 TFI 推断的光强进行初步处理，得到一系列初始重建结果$\{\hat{\mathbf{I}}(\mathbf{x},k)\mid k\in[t_{\text{key}}-T_1r_1,t_{\text{key}}+T_1r_1]\}$，其中$T_1$和$r_1$分别为长时滤波的间隔与步长。然后基于初步的重建获得更为精细的运动场，并进行基于三维时空立方体内的自回归模型的长时滤波，从而层次化地融合脉冲流中的时空信息。对于一个空间区域Ω_0而言，其对于不同时刻的初步重建影像进行融合的权重$\boldsymbol{\alpha}=\{\alpha_0(-r_1),\ldots,\alpha_0(r_1)\}$可以通过求解如下最优化问题得到：

$$\begin{aligned}&\arg\min_{\boldsymbol{\alpha}_0}\eta\,||\boldsymbol{\alpha}_0||_2^2+\\&\sum_{\mathbf{x}\in\Omega_0}\left(\hat{\mathbf{I}}(\mathbf{x},t_{\text{key}})-\sum_{k=-r_1}^{r_1}\alpha_0(r_1)\hat{\mathbf{I}}_{\text{warp}}(\mathbf{x},\hat{T}_k)\right)^2,\end{aligned}\tag{13}$$

其中η为平衡两项损失的系数，$\hat{\mathbf{I}}_{\text{warp}}(\mathbf{x},\hat{T}_k)=\hat{\mathbf{I}}(\mathbf{x}+\mathbf{w}(\mathbf{x},\hat{T}_k\,\text{vert}\,t_{\text{key}}),\hat{T}_k)$，$\hat{t}_k=t_{\text{key}}+kT_1$，基于求得的权重，可以在空间区域$\Omega_0$内实现长时窗口内的信息聚合：

$$\mathbf{I}(\mathbf{x},t_{\text{key}})=\sum_{k=-r_1}^{r_1}\alpha_0(r_1)\hat{\mathbf{I}}_{\text{warp}}(\mathbf{x},t_{\text{key}}+kT_1),\mathbf{x}\in\Omega_0.\tag{14}$$

综上所述，基于滤波的时空信息融合方法在基础脉冲成像中，旨在突破传统窗口脉冲计数法（TFP）和脉冲间隔法（TFI）方法的局限性。该研究领域的关键问题在于如何通过挖掘脉冲流中的时空上下文信息，实现噪声抑制与运动模糊之间的平衡。现有研究方法在技术路线上呈现出了不同的特点：包括沿运动方向的滤波策略、基于脉冲间隔长度相似性的滤波、基于亮度信息融合策略，以及层次化滤波架构等。这些策略均为后续的脉冲成像算法设计提供了关键启发。本部分滤波类方法主要着手处理了第 3 节所述的脉冲随机噪声和窗口长度－运动模糊权衡，通过时间窗口统计与空域滤波平滑脉冲流、抑制噪声，但在复杂运动和剧烈光照变化下对时空一致性的建模能力仍然有限。

（b）基于生物神经元机制的时空建模方法

脉冲相机模拟了生物视网膜中央凹的工作机制，其输出的脉冲流与生物视觉系统中神经元发放的神经脉冲具有相似的时序特性。在生物视觉通路中，视网膜经多层神经元处理，并依赖对时空信息

的动态整合[21]形成对场景的感知。基于这一生物学启发，研究者将脉冲相机输出的离散脉冲流视作神经元的输入刺激，构建模拟生物神经元的信息处理机制，对脉冲流进行解析并推断场景光强。

Zhu 等人[24,40]提出基于脉冲时序依赖性可塑性（Spike-Timing-Dependent Plasticity，STDP）[41-42]规则的脉冲相机影像重建算法，其中 STDP 通过根据突触前后神经元脉冲发放的时间先后调整突触连接强度，使神经元群体自适应学习脉冲流中的时序关联，从而有效编码动态信息。该重建算法设计了运动局部兴奋层、脉冲时序优化层和视觉重建层以模拟生物视网膜的信息处理过程：运动局部兴奋层结合马尔可夫随机场和图割实现动静区域划分；脉冲时序优化层基于漏电-积分-发放（Leaky-Integrate-and-Fire，LIF）神经元，根据局部运动自适应调整不应期，在静态区域抑制噪声、在动态区域保留纹理细节；视觉重建层则采用动态阈值机制融合上述结果，完成协同重建。

Zheng 等人[43]提出的 TFSTP 算法利用了基于突触短时可塑性（Short-Term Plasticity，STP）[44-45]的神经元模型。STP 机制模拟生物突触在短时间尺度上对脉冲频率变化的动态响应特性：当突触前神经元高频发放时，由于神经递质暂时耗竭，突触传递效率降低，发放频率降低后又逐渐恢复，从而使神经元对输入脉冲流的响应具有频率依赖的自适应特性，可在高速运动场景中对快速变化的脉冲序列产生适应性衰减、对缓慢变化的脉冲保持稳定响应，实现对时空信息的自适应编码。

TFSTP 包括 STP 动态建模层、运动区域优化层和协同重建层：其中，STP 动态建模层通过神经递质剩余量 R 和释放概率 u 的动态变化隐式估计脉冲发放频率；运动区域优化层基于 R 值的帧间波动检测运动像素，并通过局部连接的 LIF 神经元网络优化运动区域轮廓；协同重建层对静态区域采用 STP 估计结果以保留细节，对运动区域采用小窗口 TFP 方法以抑制动态模糊并融合得到重建结果。随后提出的 TFMDSTP[46]则在 STP 基础上根据其动态响应区分动静区域，使用两组不同参数的 STP 模型分别重建静态与动态区域。

综上所述，基于生物神经元机制的时空建模方法，通过模拟生物视觉系统的 STDP、STP 等机制，结合 LIF 神经元构建多层架构，动态编码脉冲流时空关联，实现动静区域区分以处理场景中的运动，从而实现了脉冲成像中的时空信息聚合。

本部分基于生物神经元的成像方法侧重应对第 3 节中的动静区域区分与长时上下文利用不足等问题，利用 LIF、STDP、STP 等模型在时间上自适应积累脉冲并区分动静场景，但重建质量仍受神经元模型逼真度和参数设置的限制。

（c）基于端到端深度学习的时空信息融合方法

随着深度学习在成像领域的广泛应用，研究者们开始探索利用端到端神经网络直接从脉冲流中学习时空特征，以实现高质量的脉冲成像。这类方法可以一定程度上摆脱传统手工设计特征的局限，通过深度神经网络提取脉冲流中的时空上下文，最终直接映射至待重建时刻的场景光强。

与基于滤波的方法和基于生物神经元的方法类似，这类方法也要重点考虑噪声的抑制与运动的对齐，不同的方法在这两个方面进行了不同角度的尝试。除此之外，深度学习往往需要基于大量数据进行训练，而在脉冲成像领域中，高质量的脉冲流与清晰影像的配对数据获取难度很大，一般通过脉冲模拟器基于视频数据集进行生成，这限制了端到端深度学习方法的泛化能力。为此，一些方法[40,47]研究了基于自监督学习的脉冲成像算法。基于深度学习的脉冲影像重建方法的总结如表 2 所示。

最早期的深度学习脉冲影像重建方法为 Spk2ImgNet[20]。如图 6 所示，该方法提出了“短时窗口特征提取－长时运动对齐－多时刻特征融合”的端到端框架：首先在较短时间窗口内提取局部特征，在抑制大幅运动引起模糊的同时，通过聚合有限时域信息降低噪声；随后以待重建时刻及其若干邻近时刻为中心提取多组短时特征，并通过时空对齐模块统一到目标时刻；最后对对齐后的多时刻特征进行融合，得到清晰影像，为后续基于深度学习的脉冲成像算法提供了典型范式。不同的脉冲成像算法在特征提取与时空对齐上采用了各异的设计。以 Spk2ImgNet 为例，其在特征提取阶段引入尺度时序滤波与空间自适应选择机制，可根据运动速度和光照条件动态调整时序尺度；在特征对齐阶段采用基于可变形卷积的金字塔对齐模块[48-50]实现由粗到细的时域对齐；在特征融合阶段则基于可靠性预测进行加权融合，有效减弱运动模糊和对齐失效对重建结果的影响。

在后续研究中，不同工作从多种思路对深度学习脉冲成像进行了改进。Chen 等人[51]提出首个自监督脉冲成像方法 SSML，旨在缓解监督方法对大规模合成标注数据的依赖：一方面，合成数据与真

表 2 基础脉冲成像算法中基于端到端深度学习的时空信息融合方法

方法	发表信息	亮点	算法核心
Spk2ImgNet[20]	CVPR 2021	首个深度学习方法	长短时长窗口特征提取，金字塔可变形卷积时域对齐
SSML[51]	IJCAI 2022	首个自监督深度学习方法	自适应运动推理，盲点网络自监督学习
Chen 等[52]	AAAI 2023	自监督学习运动-影像联合估计	影像与运动联合自监督估计，重渲染损失
WGSE[53]	AAAI 2023	基于小波变换的特征提取	基于多尺度小波分解的脉冲特征提取
RSIR[54]	MM 2023	实现对不同光照的自适应处理	自适应脉冲密度调整，小波域循环对齐
Hu 等[55]	ECCV 2024	针对暗光场景进行成像	基于全局脉冲积分周期的光照表征，双向循环时域对齐
SSIR[56]	TCSVT 2024	首个深度脉冲神经网络方法	基于自适应神经元的脉冲特征提取，重参数化的输出层
STIR[57]	NeurIPS 2024	高效轻量化重建	显示-隐式联合脉冲特征提取，运动-光强交互估计
SpikeODE[58]	TCSVT 2024	基于神经常微分方程利用脉冲连续性	将场景的连续动态过程通过神经常微分方程嵌入到隐状态中
BSF[33]	CVPR 2024	针对脉冲随机波动性进行分析与处理	基于多尺度脉冲发放时间差分的特征提取，基于搜索策略的多粒度对齐
Spk2ImgMamba[59]	WACV 2025	引入了多尺度状态空间模型	提出了轻量多尺度状态更新与渐进式重建架构提升时空结构表征
Spike-CLIP[60]	AAAI 2025	将文本-图像对齐引入脉冲成像	基于文本-图像对齐的多模态监督训练

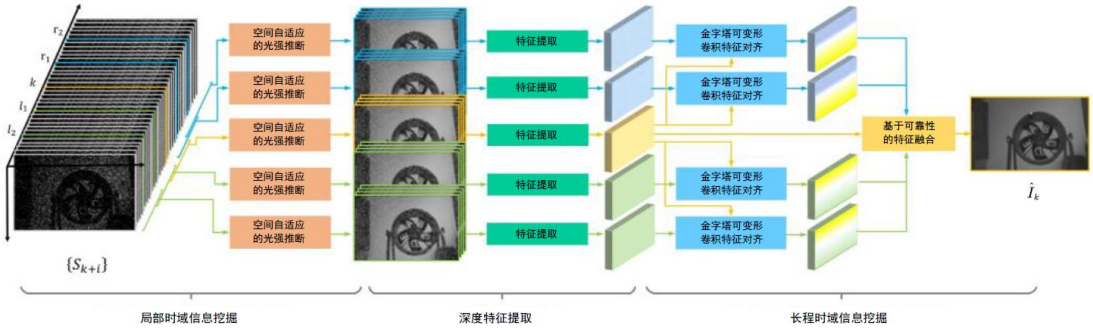

图 6 Spk2ImgNet[20]的网络架构

实脉冲流在噪声分布上存在差异，另一方面，真实场景的脉冲－清晰影像配对获取成本极高，均限制了模型的泛化能力。SSML 采用盲点网络与非盲点网络的互监督机制实现无标注训练，其中盲点网络以 TFP、TFI 等传统成像结果为伪标签，结合盲点约束[61]学习潜在低噪影像；非盲点网络则集成非局部注意力等模块挖掘长时时空关联，二者通过知识蒸馏相互促进，得到清晰重建结果。

在此基础上，Chen 等人[52]进一步提出自监督联合动态场景重建与光流估计框架，将两项任务联合训练，形成任务的协同闭环；同时引入基于物理约束的逆渲染自监督机制，通过利用光流插值生成中间帧并模拟长曝光模糊，与脉冲流长时聚合影像进行对比构建模糊一致性损失，从而增强重建结果与脉冲成像物理特性的匹配程度。

Zhang 等人[53]提出的 WGSE 将小波变换引入脉冲成像，通过多尺度离散小波变换将脉冲流分解为反映瞬态变化的高频系数和体现整体趋势的低频系数，再利用时序残差卷积网络对各尺度系数进行优化，最后通过逆小波变换重构出兼具时频特征的脉冲流表征，从而更准确地刻画离散脉冲的时序特性。Zhu 等人[54]提出的 RSIR 方法面向高、低光照等不同光照条件下的通用脉冲成像问题，分析了不同光照下脉冲流噪声分布的差异，并通过自适应脉冲变换模块动态调整处理窗口长度，同时结合基于频域分解的循环时间融合进一步挖掘时域上下文信息。Hu 等人[55]针对低光环境下脉冲流稀疏导致的成像困难，提出基于全局脉冲发放间隔的鲁棒光强表征，并构建双向循环网络框架对时间上下文进行融合，以提升低光条件下的重建质量。

Zhao 等人[56]提出的 SSIR 首次将深度脉冲神经网络用于脉冲成像，利用脉冲神经网络事件驱动、适配时序信号的特性，设计了漏电参数自适应的脉冲残差模块，并基于重参数化技术[62]构建渐进式滤波模块，在影像重建中实现了重建精度与效率的平衡。Fan 等人[57]提出轻量化 STIR 框架，采用混合脉冲嵌入表征融合显式初步重建与隐式特征，在顶层通过对称交互注意力块建模中间帧与时空上下文的双向关联，在底层结合多运动场估计补偿细节，并由时空交互解码器从粗到细联合执行对齐与滤波，实现高效的一体化特征提取与运动对齐。

Yang 等人[58]提出 SpikeODE，将神经常微分方

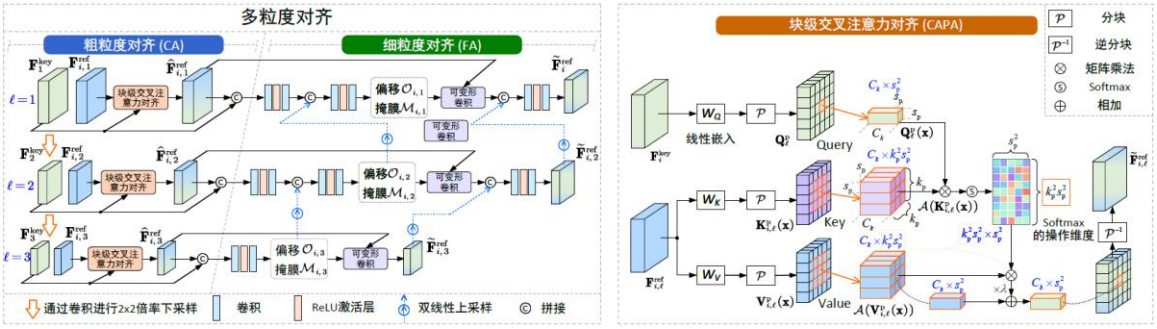

图 7 BSF 网络[33]中提出的多粒度对齐

程引入脉冲成像，通过隐状态建模恢复场景的连续动态过程，以挖掘时间相关性。该方法在每个时间戳以观测脉冲为条件进行前向和后向双向推理，融合双向时间上下文，从而在连续动态建模框架下进一步提升脉冲成像质量。Yin 等人提出 Spk2ImgMamba[59]，将多尺度状态空间模型[63]引入脉冲成像，以紧凑骨干网络同时建模脉冲流中的长短时上下文信息。该方法通过多尺度状态更新与渐进式重建提升对时空结构的表征能力，在保持计算效率的同时进一步改善了重建质量。

上述工作从不同角度优化了脉冲相机的基础成像问题，但对脉冲流噪声特性的专门分析仍相对有限。Zhao 等人[33]通过分析脉冲流中的噪声，揭示了脉冲表示光强的“波动性”特征，并据此设计了 BSF 网络以实现高质量成像。这里的脉冲波动性指的是：即便光照强度保持稳定，脉冲发放间隔仍会随时间上下波动，这主要由光子的泊松统计、电路热噪声和量化效应等因素共同造成，从而使单次脉冲间隔难以直接反映真实光强。Zhao 等人进一步发现，脉冲发放时间差分（Differential of Spike Firing Time，DSFT）的倒数可以构造对光强的无偏估计；据此提出多阶 DSFT 融合模块，将 DSFT 扩展到更高阶，也即涵盖多个脉冲间隔，通过融合不同阶数的 DSFT 特征，在保留低阶 DSFT 动态细节的同时利用高阶 DSFT 在长时间尺度上的稳定性抑制噪声，有效减弱脉冲波动的影响。针对时间对齐问题，如图 7 所示，BSF 网络设计了多粒度对齐模块：在以往金字塔对齐的基础上，引入基于搜索策略的块级交叉注意力，以拟重建时刻特征为查询、参考时刻特征为键和值，在局部范围内匹配对应区域，利用局部光强的相对稳定性降低脉冲波动造成的匹配误差；在此基础上，细粒度对齐通过可变形卷积预测像素级偏移和掩码，进一步校正运动偏移，两级对齐协同提升了特征对齐的鲁棒性。

表 3 展示了主要的基础重建方法在 REDS-SCIR 数据集[33]上的性能表现，以及其中的深度学习方法的参数量、计算复杂度与运行时间。其中参数η是 REDS-SCIR 数据集中与场景亮度相关的系数，η越大，场景越亮。表中的后五行为深度学习方法，均在 REDS-SCIR 上进行了重新训练，其中 Spk2ImgNet、WGSE 与 BSF 采用了 BSF[33]中的训练方法；SSML 与 STIR 采用了其原文中的训练方法。表中的数据通过 NVIDIA RTX 3090 GPU 在分辨率为250×400的数据上推理得到，其中计算量通过乘加运算（MACs）体现。

考虑到传统监督方法受合成与真实数据域差距的影响、自监督方法依赖低质量伪标签，Chen 等人[60]提出了 SpikeCLIP，将 CLIP 模型[64]的文本－图像对齐能力引入到脉冲成像中，构建了基于文本描述和非配对高质量数据的新型监督范式。其突破了传统重建方法对成对标注数据的依赖，利用场景的文本类别信息和高质量影像特征指导网络学习，提升了低光场景下的纹理恢复与亮度平衡能力。本部分深度学习类成像方法着重处理了第 3 节提及的复杂噪声建模、长时空依赖挖掘以及低光/高动态范围成像等挑战，通过端到端特征提取、时域对齐和多尺度融合显著提升困难场景下的重建质量，同时也带来了对大规模数据和算力资源的更高需求。

相比事件相机重建方法通常需要通过事件积分或与帧式相机联合补全静态背景，上述脉冲成像算法直接在脉冲流上同时恢复静态纹理和高速运动细节，充分利用了脉冲相机的静态信息保留能力，在弱纹理静态区域成像质量上具有明显优势。

4.2 脉冲超分辨率成像算法

在基础脉冲成像实现灰度重建的基础上，脉冲超分辨率成像借助脉冲流的高时间分辨率与场景运动特性，从低空间分辨率脉冲流中重建高分辨率

表 3 REDS-SCIR 数据集上不同脉冲成像方法的对比

成像方法	$\eta=1.00$			$\eta=0.75$			$\eta=0.50$			参数量（M）	MACs（G）	推理时间（ms）
	PSNR ↑	SSIM ↑	LPIPS ↑	PSNR ↑	SSIM ↑	LPIPS ↑	PSNR ↑	SSIM ↑	LPIPS ↑			
TFP[18]	27.27	0.711	0.265	26.73	0.669	0.300	25.62	0.581	0.370	—	—	—
TFI[18]	23.55	0.634	0.329	24.77	0.673	0.293	26.77	0.713	0.249	—	—	—
TFSTP[43]	20.35	0.678	0.270	19.62	0.685	0.252	21.10	0.707	0.247	—	—	—
MAHTF[19]	29.57	0.879	0.112	30.07	0.884	0.113	29.65	0.869	0.136	—	—	—
SSML[51]	33.94	0.923	0.075	33.27	0.909	0.088	32.01	0.883	0.116	2.385	386.02	408.3
Spk2ImgNet[20]	39.16	0.966	0.024	38.27	0.958	0.032	36.59	0.940	0.051	3.904	1016.38	167.6
WGSE[53]	38.97	0.964	0.027	38.23	0.957	0.034	36.75	0.940	0.049	3.806	425.22	88.7
STIR[57]	37.28	0.946	0.052	36.47	0.937	0.061	34.95	0.916	0.081	5.108	46.60	66.6
BSF[33]	39.76	0.970	0.021	39.09	0.964	0.027	37.76	0.951	0.040	2.477	726.89	207.5

动态场景。由于相机与场景的相对运动，同一像素的脉冲会记录场景中不同位置的光子信息，每个脉冲对应特定时刻和空间点的光强采样，从这类脉冲流中恢复高分辨率细节的过程被称为脉冲超分辨率重建。这一过程不仅需解决光照强度解码、噪声抑制，还需要精准估计运动以建立脉冲与高分辨率像素的对应，同时利用时间连续性挖掘空间细节，在继承脉冲解码需求的同时，新增了运动－空间关联建模及时间信息向空间细节转化的目标。

Zhao 等人[25]提出的运动引导超分辨率（MGSR）框架，首次实现了从低分辨率连续脉冲流直接重建高分辨率动态场景，其超分辨成像模型体现了对脉冲与光强关系的深度解析与量化建模。该模型的核心在于：利用相机与场景的相对运动，将同一传感器像素因运动记录的不同空间位置信息转化为超分依据。具体而言，该模型通过一系列公式详细刻画了脉冲与光强的内在关联。首先，基于脉冲相机的“积分－发放”机制，对于原始分辨率下像素 p 处的一个脉冲 s_0，其累积起始和结束时刻分别为 t_s 和 t_e，该累积过程可以被表示为：

$$\theta=\int_{\Omega_p}\int_{t_s}^{t_e}\alpha\cdot\mathbf{I}(\mathbf{x},t)\ \mathrm{d}t\,\mathrm{d}\mathbf{x}, \tag{15}$$

其中 θ 为脉冲的发放阈值，Ω_p 为像素 p 所处的区域，α 为光电转化效率，$\mathbf{I}$ 为场景的光强，$\mathbf{x}$ 和 t 分别为空间和时间坐标。如果只关注拟重建时刻 t_{key} 的光强，基于该时刻到各个时刻的运动场 $\mathbf{w}(\mathbf{x},t|t_{\mathrm{key}})$，上式可以被重写为：

$$\theta=\int_{\Omega_p}\int_{t_s}^{t_e}\alpha\cdot\mathbf{I}(\mathbf{x},t_{\mathrm{key}})\cdot\mathcal{M}_p(\mathbf{x},t)\ \mathrm{d}t\,\mathrm{d}\mathbf{x}, \tag{16}$$

其中二进制掩码 $\mathcal{M}_p$ 指示了坐标 $\mathbf{x}$ 处的高分辨率像素是否在 t 时刻落入原始分辨率下像素 p 的区域：

$$\mathcal{M}_p(\mathbf{x},t)=\begin{cases}1, \mathbf{x}+\mathbf{w}(\mathbf{x},t\,\mathrm{vert}\,t_{\mathrm{key}})\in\Omega_p,\\ 0, \mathbf{x}+\mathbf{w}(\mathbf{x},t\,\mathrm{vert}\,t_{\mathrm{key}})\notin\Omega_p.\end{cases} \tag{17}$$

如果设贡献权重 $\mathcal{W}_p(\mathbf{x})$ 为拟重建时刻的高分辨率像素 $\mathbf{x}$ 对上述原始分辨率坐标 p 处脉冲 s_0 的贡献：

$$\mathcal{W}_p(\mathbf{x})=\int_{t_s}^{t_e}\mathcal{M}_p(\mathbf{x},t)\ \mathrm{d}t. \tag{18}$$

则原成像公式可以被表示为：

$$\theta=\int_{\Omega_p}\int_{t_s}^{t_e}\alpha\cdot\mathbf{I}(\mathbf{x},t_{\mathrm{key}})\cdot\mathcal{W}_p(\mathbf{x},t)\ \mathrm{d}\mathbf{x}, \tag{19}$$

上述模型通过建立光强与每个脉冲的映射关系，提出使用贡献权重来量化高分辨率像素对脉冲的贡献度，即每个脉冲对应场景中特定时空点的光强积累，而贡献权重则由运动轨迹与脉冲周期共同决定。从这种物理模型出发，基于各个脉冲提供的约束对拟重建的高分辨率帧进行求解，可以使得脉冲流的丰富时间上下文得以充分利用。

Xiang 等人[65]从端到端深度学习的思路出发，提出了脉冲相机的超分辨率成像网络。该方法首先利用一维卷积提取流内时间特征，捕捉脉冲流的高时间分辨率信息；其次借助光流模块提取流间时空依赖，建模中间脉冲流与相邻流的关联；最后通过基于脉冲的迭代投影模块进行联合特征学习，将流内与流间特征分别映射至高分辨率空间，计算残差并投影到流间多特征上，最终实现端到端的脉冲超分辨率成像。

Zhao 等人[66]提出的 SpikeSR-Net 融合了成像模型驱动与端到端深度学习这两方面的特点。该网络基于脉冲相机超分辨率观测模型，在 Spk2ImgNet 网络的整体架构基础上设计了深度展开的神经网络。在特征提取过程中，其设计了像素自适应的脉冲编码器，通过卷积神经网络生成了像素自适应的卷积核，以提取鲁棒的光强特征。然后，其基于脉冲超分辨率观测模型求解的优化过程，通过可变形卷积网络迭代执行了时域上下文利用过程中的梯度下降与近端优化，并且基于双向循环网络进一步融合了长时的相关性，实现脉冲超分辨率成像。

Wang 等人[67]提出的 Spk2SRImgNet 网络，针对脉冲流因泊松效应、量化效应及电路热噪声产生的波动问题，设计了运动对齐协同滤波模块。该网络先通过脉冲间隔重建将二进制脉冲流转换为初步光强信息，再经编码器提取特征。然后，运动对齐协同滤波模块先估计运动偏移量，通过可变形卷积实现运动对齐，然后在变换域对堆叠的相似特征进行协同滤波生成细化特征，经逆运动对齐映射回原位置后再进行基于可变形卷积的运动对齐，而后与初始特征聚合。算法最终融合对齐后的不同时刻特征，并通过像素重排输出高分辨率影像。

本小节的超分辨率方法主要针对第 3 节中“空间分辨率受限、难以兼顾高时间采样与高空间细节”的问题，通过多尺度重建和时域信息向空域细节的转换，将脉冲相机的时间分辨率优势转化为更高的空间分辨率。在超分辨成像中，相较于事件相机依赖于运动边缘获取亚像素信息，脉冲相机还能进一步对静态纹理进行连续记录，从而挖掘其中的高频细节。

4.3 彩色脉冲成像算法

在基础脉冲成像实现灰度重建的基础上，彩色脉冲成像通过相机中彩色滤光阵列（CFA）实现彩色动态场景记录与恢复。和传统相机类似，其依赖 CFA 对不同波段光的选择性透过（如 Bayer 模式），每个像素仅响应红、绿、蓝（R、G、B）中一种，形成单通道脉冲流，从该采样信号中恢复完整彩色影像即去马赛克成像。这种单通道采样使直接获取完整彩色信息成为新的挑战：需要进行基础脉冲成像的光照强度恢复，补全缺失颜色通道信息，同时兼顾脉冲流高时间分辨率，避免运动模糊与色彩混叠。因此，彩色脉冲成像在继承基础算法脉冲解码需求的同时，新增加了跨通道相关性建模与时空一致性维护目标。

3DRI[26]是彩色脉冲成像的早期方法，其先通过脉冲间隔预处理将彩色脉冲流转为单通道影像信号，按 Bayer 模式分解为 R、G、B 三个含缺失像素的通道，再分别用光流估计获取运动轨迹，沿轨迹对邻近帧像素滤波与线性插值，以既利用时序相关性抑制量化噪声，又避免固定窗口滤波导致的模糊。因绿色通道采样更丰富，所以以插值后的 G 通道为引导，对 R、B 通道上采样，再计算 R、B 通道与 G 通道的残差并对残差执行沿运动的 3D 插值，最终融合得完整彩色影像。

CSpkNet[68]是早期基于深度学习的彩色脉冲成像方法，通过三阶段架构实现从 Bayer 模式脉冲流到彩色影像的重建。先由光推断模块将脉冲流分解为 R、G1、G2、B 子序列，再通过通道注意力调整时间窗口内各帧权重，抑制量化噪声导致的光强估计误差。其中运动引导滤波模块基于 RAFT 光流网络[69]估计运动轨迹，按 Bayer 模式对齐像素，结合颜色融合注意力学习滤波权重以利用跨通道相关性，最终通过残差模块网络实现彩色影像重建。

SJDD-Net[70]针对彩色脉冲流中的复杂噪声，设计了联合脉冲流去马赛克与去噪框架。参考优化问题迭代求解策略，通过二次半分裂（HQS）将联合问题拆分为保真项与先验项两个子问题。随后，将迭代过程展开成一个多阶段神经网络。每阶段中，保真项对应颜色脉冲表征模块，通过多尺度 3D 编码器与可变形卷积获取运动对齐特征，为像素补全打基础；先验项由颜色相关先验模块处理，结合脉冲噪声估计器输出的噪声特征，利用颜色通道相关性修复缺失像素并抑制噪声，两者迭代协同优化，最终构建清晰彩色影像。

CSCSR[71]网络是一种彩色脉冲联合去马赛克与超分辨率成像的方案，通过端到端架构实现从低分辨率脉冲流到高分辨率彩色影像的直接重建。在局部与全局联合时空特征提取基础上，估计亚像素级运动，根据运动轨迹与颜色阵列布局在时域搜索相关像素，将时间分辨率转化为空间分辨率。Yang 等人[72]提出实数据驱动的彩色脉冲成像方法，面向彩色脉冲相机的高速彩色 HDR 视频重建，在无需成对标注的条件下直接以实采马赛克脉冲流训练网络。该方法基于短时累积构建自监督去噪模块抑制脉冲噪声，并通过渐进式时域配准将多帧短时影像沿运动轨迹对齐以模拟无模糊长曝光，从而实现基于马赛克彩色脉冲流的彩色 HDR 视频重建。

针对彩色脉冲成像中合成与真实数据的域差距问题，Dong 等人[73]提出基于自监督学习的彩色重建方法。首先通过联合运动估计利用颜色通道运动一致性，融合各通道光流提升轨迹准确性，沿运动轨迹收集对应颜色的时间相邻像素填补缺失像素，对无像素位置采用空间插值，初步解决通道像素缺失与运动模糊。在此基础上，以初步重建结果为伪标签，通过掩码 3D 部分卷积分离颜色通道并提取时空特征，并采用带时间盲点的时域掩码卷积抑制时间相关量化噪声，两者分支特征融合后经残差块处理，最终得清晰影像。

为缓解深度学习方法对数据的依赖，Dong 等

人[74]提出无需预训练的彩色脉冲成像方法。基于脉冲流超高时间分辨率，用滑动时间窗口采样片段构建“附近时刻”与“中间时刻”自监督对，通过迭代优化网络参数实现动态调整。并利用光流估计运动轨迹，沿运动方向收集时间相邻像素生成运动引导伪标签作为监督，填补颜色通道缺失像素，基于时间上下文进一步提升成像质量。相比仅对亮度变化敏感的事件相机，脉冲相机在彩色场景下可以同时记录静止区域的颜色细节和运动区域的边缘动态，因此在彩色纹理保真度和伪影抑制方面更具有优势。

4.4 脉冲高动态成像算法

在彩色成像的基础上，HDR 成像进一步关注复杂光照下亮暗细节的共同恢复。基础脉冲成像可实现灰度重建，但脉冲相机应对高动态范围（HDR）场景时，难以兼顾高光与低光细节。受单阈值积分限制，强光区脉冲易丢失精细信息，弱光区因脉冲稀疏致而导致光强估计误差大。脉冲高动态成像算法通过优化脉冲编码与重建拓展动态范围，Chang 等人[75]的滚动混合位脉冲机制、Zhu 等人[76]的多级脉冲相机模型是典型方案。

Chang 等人[75]针对脉冲相机高动态与高帧率的矛盾，也即单比特脉冲低冗余但动态范围有限，多比特脉冲增加数据量的问题，提出滚动混合位（RMB）脉冲机制与新相机模型，其中“滚动混合位”指在逐行读出的单比特脉冲流中周期性插入少量多比特脉冲，以在基本不增加数据量的前提下提升有效比特深度。如在 500×500 分辨率帧中，每次仅读 4 行 3 比特脉冲，其余行读出单比特。这样可以使得总数据量较纯单比特仅增不足 2%，同时保留了多比特扩动态能力。为解决滚动读取可能的形状畸变，模型基于脉冲发放时间差分（DSFT）估计运动信息，结合光流映射对齐运动；随后，通过跨比特注意力块融合单/多比特特征，并通过跨时间注意力块学习时序权重掩码以进行时域融合，最终实现高动态、高帧率影像重建。

Zhu 等人[76]设计了多级脉冲相机模型及对应重建方案。模型核心是两级缓冲机制与指数编码的结合：一级缓冲为积分器，累积光子达阈值触发脉冲并重置，精细记录光子到达；二级为脉冲触发计数器，累计单次读取间隔内未读出脉冲，旨在解决强光区脉冲信息记录不完整的问题。重建时，基于状态空间模型[63, 77]，该方法首先通过时空选择性扫描提取时序关联，再结合空间关联，并利用金字塔自适应滤波与对齐模块实现从粗到细的特征对齐，从而高效利用脉冲流的时空信息。

HDR 脉冲成像方法主要面向第 3 节中“高动态范围与低光场景下易饱和、易欠曝”的问题，利用脉冲在强弱光下的自适应发放特性设计时域聚合与动态范围扩展策略，实现对亮暗区域细节的联合重建。相较于事件相机，脉冲相机在进行 HDR 成像时更能兼顾静态区域高光与暗区的细节保留。

4.5 特殊场景脉冲成像算法

在突破脉冲相机动态范围、实现高动态成像的基础上，脉冲成像技术在更具挑战性的特殊场景中展现出了独特应用价值。研究者借脉冲流时序优势建模场景动态关联，从冗余/不完整观测中提关键信息，拓展了其在精密测量、科学观测等领域的应用。

高速相变过程（如材料凝固、水结冰）发生快且起始时间不定，传统相机难以长时高帧率记录，脉冲相机则因其高时间分辨率成为理想工具。Xiong 等人[27]提出自监督脉冲流重建算法，其核心是双网络自监督架构：非盲点网络以传统脉冲成像结果为伪标签，结合盲点约束学习低噪声特征；盲点网络挖掘长时时空关联提深层动态信息，二者通过知识蒸馏互相监督，最终由非盲点网络输出高质量结果，无需标注数据，恢复了相变瞬态细节并实现了降噪。

高速微流控液滴系统中，液滴频率、速度、尺寸等参数十分关键，但 1-10kHz 的高速特性让传统检测面临挑战：一维光学探测缺少空间信息，影像法受数据吞吐量、运动模糊限制。Xiong 等人[78]提出脉冲流实时液滴参数评估算法，直接从脉冲流提取参数，方法分三步实现：首先根据频率评估选择感兴趣区域，对脉冲流时域累积后用傅里叶变换算功率谱密度，识别主峰并拟合频率；进而进行速度评估，对于两已知间距检测区，借助互相关析相位差，结合频率计算速度；最后进行尺寸评估，基于脉冲间隔统计特性，设置阈值提取边缘脉冲间隔变化，结合速度，转换时间间隔为空间尺寸，并基于多像素统计融合降低噪声。

动态场景中前景遮挡（如栅栏、网格等）阻碍了背景的感知，传统影像去遮挡法依赖多相机阵列或多视角影像，受视角不足、运动模糊等限制。Zhang 等人[79]提出脉冲相机连续多视角去遮挡方法 SpkOccNet，借助其高时间分辨率与连续记录特性，从单相机脉冲流中重建被遮挡背景。其算法的核心是挖掘脉冲流连续时空关联：先以长短窗口特征提

取器分段处理脉冲流，其中辅助段用密集窗口保时序细节，中心段兼用长窗口（获整体结构）与密集窗口（捕捉动态），再经 U 型网络提取特征；随后跨视角互注意力融合模块融合中心段与辅助段特征，借视角互补补充遮挡信息，最终经通道注意力优化结果，实现高效去遮挡成像。

4.6 脉冲三维成像算法

在脉冲成像技术演进中，基础成像、超分辨成像等二维成像聚焦从脉冲流恢复单视角光强信息，构建场景二维光强表征；三维成像则突破维度限制，整合多视角时空信息，实现场景三维光照强度完整刻画。近年来，神经辐射场（Neural Radiance Field，NeRF[28]与 3D 高斯泼溅（3D Gaussian Splatting，3DGS）[29]凭借其三维表征能力成为研究热点：NeRF 通过隐式神经辐射场建模实现高质量新视角生成，3DGS 以显式高斯椭球基元表示场景实现实时高保真渲染，两者均突破了传统基于影像的三维重建在动态场景捕捉与渲染效率上的瓶颈。

脉冲相机的超高时间分辨率与连续记录特性为三维成像提供优势：其输出脉冲流蕴含密集时空动态信息，既能捕捉高速运动细节，又能通过脉冲发放记录光强连续变化，为解决传统相机在高速动态场景中因帧率不足导致的运动模糊、信息丢失等问题提供新思路。二维重建是脉冲三维成像基础：前者通过解析脉冲流恢复单个时刻光强分布，为三维提供关键单视角光强先验；后者在此基础上，利用脉冲流时序连续性与多视角关联，结合 NeRF 或 3DGS 的三维建模管线，实现从 “二维光强序列” 到 “三维场景结构” 的提升。

SpikeNeRF[80]首次实现从脉冲流学习基于 NeRF 的隐式三维场景表示，突破传统 NeRF 对高质量传统影像的依赖。其核心是构建适配脉冲特性的建模框架：引入积分－发放（IF）神经元层模拟脉冲相机“积分－发放”机制，通过非均匀矩阵模拟像素级阈值变化，捕捉散粒噪声、暗电流噪声等非理想特性，提升对脉冲流的建模准确性。为解决脉冲流噪声干扰，该方法设计长期脉冲渲染损失，通过长时间窗口内生成与真实脉冲流的统计特性比对实现自监督学习，结合替代梯度法和沿时间反向的传播完成端到端优化。

Zhang 等人[81]首次将 3DGS 与脉冲相机结合，提出 SpikeGS 框架，解决传统相机在高速动态场景中因运动模糊导致的 3D 成像质量下降问题。该框架核心是构建适配脉冲流特性的 3DGS 训练机制：一方面设计“瞬时成像”与“类曝光成像”双约束，瞬时成像通过基于盲点网络架构的轻量级脉冲瞬时映射网络从局部脉冲流推断瞬时清晰影像，为 3DGS 提供单个时刻精细监督；类曝光成像通过累积脉冲流与 3DGS 连续渲染结果的均值构建约束，利用脉冲时间连续性弥补高速运动中视角信息缺失，模拟长曝光优化动态场景全局结构。另一方面结合瞬时损失和类曝光损失优化 3D 高斯基元参数，兼顾细节纹理与动态一致性。

Yu 等人[82]同期也提出了同名的基于 3DGS 的脉冲流三维成像框架 SpikeGS，旨在解决传统 NeRF 在脉冲相机三维成像中计算复杂度高、对噪声鲁棒性不足等问题，实现从纯脉冲流高效学习三维高斯场。该框架核心是构建可微分脉冲流渲染管线，含噪声嵌入、脉冲神经元层和脉冲渲染损失三部分设计：通过响应非均匀性矩阵模拟脉冲流各类噪声，利用积分-发放机制的脉冲神经元层将 3DGS 渲染光强转换为脉冲流，通过采用结合了 L1 与 D-SSIM 的混合损失函数，直接对齐生成与真实脉冲流，无需依赖中间影像重建。该方法较 SpikeNeRF 实现明显速度提升。

Guo 等人[83]提出首个将 Bayer 模式彩色脉冲流与 3DGS 结合的三维成像框架，针对高速运动的彩色脉冲相机场景，解决传统 3DGS 对清晰影像的依赖及现有脉冲三维成像方法在彩色场景和动态适应性上的不足。如图 8 所示，该框架核心是适配彩色脉冲流的处理机制与优化策略：一方面设计累积光栅化技术，通过长时间累积脉冲流时序信息减少时间混叠和运动模糊，结合拜尔滤镜分通道处理，从彩色脉冲流稳定提取纹理细节；另一方面引入区间监督机制，利用脉冲间隔信息初始化点云和相机姿态，加速高斯泼溅收敛，同时动态调整累积损失与区间损失权重，平衡纹理恢复与几何精度。此外，框架首次支持彩色脉冲相机三维成像，通过模拟脉冲相机“积分－发放”物理过程，构建从拜尔模式脉冲流到三维高斯场的完整映射，为高速彩色动态场景实时三维感知提供新方案。

为解决脉冲相机三维成像中累积误差大、对高质量输入影像依赖强的问题，Chen 等人[49]提出协同优化框架 USP-Gaussian，将脉冲影像重建、姿态校正和 3DGS 整合为端到端系统。该框架一方面设计长－短脉冲流互补输入的脉冲影像重建网络，从脉冲流恢复高质量影像，为 3DGS 提供精细初始输入，重建网络采用基于运动重模糊损失和多重新模

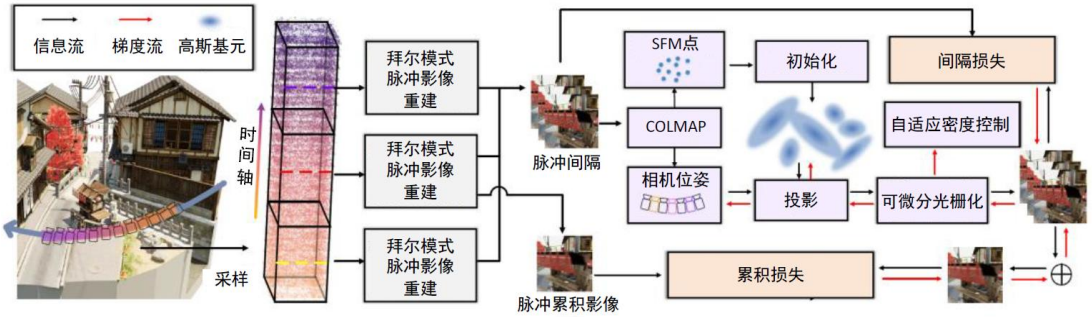

图 8　Guo 等人[83]提出的脉冲三维成像算法

糊损失的训练策略，确保重建影像在像素层面准确且保留场景结构信息；另一方面 3DGS 利用多视角一致性，通过类似运动重模糊损失优化高斯基元参数，并将重建结果反馈给重建网络约束其输出，使其更符合真实场景物理特性。此外，框架含姿态优化模型，通过李代数插值优化相机姿态，生成更精确的相机轨迹。通过联合损失函数，USP-Gaussian 实现脉冲影像重建与 3DGS 的协同优化，打破传统级联框架中各模块独立优化的局限，有效提升高速动态场景下三维重建的精度与鲁棒性。综合来看，NeRF 结合脉冲相机更侧重利用连续时间信息提升新视图重建精度，但渲染效率较低，且对场景静态假设依赖较强；3DGS 则在保证较高成像质量的同时，显著降低渲染开销，更适合实时或低动态场景。整体而言，基于脉冲流的 NeRF / 3DGS 方法在快速大幅运动和在线更新方面的动态适应性仍有提升空间。

三维脉冲成像方法则针对第 3 节提出的“深度/几何信息获取不足”这一挑战，通过视差估计、体渲染或多视几何建模，将脉冲流的时序与多视信息转化为稳定的几何与深度表征。相较于更侧重动态轮廓的事件相机，脉冲相机可结合静态纹理做更稳定的几何和外观估计。

综上所示，本节围绕脉冲相机成像算法展开，核心是解决脉冲流到清晰影像的解码问题，从基础灰度重建逐步拓展至超分辨率、彩色、高动态、特殊场景及三维成像等细分方向，形成覆盖多维度需求的技术体系。基础脉冲成像算法作为核心起点，从早期依赖单像素统计的 TFP、TFI 方法，逐步发展为三类主流技术路线：基于滤波的方法通过运动估计与时空滤波平衡噪声抑制与运动模糊，基于生物神经元机制的方法模拟神经元实现光强求解，基于深度学习的方法通过端到端网络提取时空特征。

在此基础上，各细分方向进一步突破脉冲相机性能边界：超分辨率成像借助运动关联将脉冲高时间分辨率转化为空间细节；彩色成像通过彩色滤光阵列处理与跨通道融合补全缺失颜色信息；高动态成像通过新型脉冲编码与重建策略拓展光强记录范围；特殊场景成像针对高速相变、微流控等场景定制时序特征挖掘方法；三维成像则结合 NeRF、3D 高斯泼溅等范式，将脉冲流时空信息转化为三维场景结构。整体而言，脉冲相机成像算法的演进，本质是对脉冲流时空关联的深度挖掘与利用，既解决了脉冲信号解析的核心问题，也为后续多传感器融合奠定单设备成像能力基础，支撑脉冲视觉技术从基础信号处理向复杂场景应用落地推进。

本节主要聚焦于脉冲相机从脉冲信号到影像的成像过程，在实际应用中，许多基于脉冲的视觉任务，如光流估计、目标检测等，直接在二值脉冲流上进行特征提取，进而执行对应的任务。虽然这些方法并不显式重建影像，但现有成像算法在噪声抑制、时域上下文建模以及静态 / 动态信息分离等方面的设计，仍为下游特征表示的构建提供了重要启示。例如 SpikeTransformer[17]和 HiST-SFlow[13]分别利用多段窗口的脉冲，经过噪声抑制后，融合长时窗口内的信息，对深度和光流进行了估计；SpiReco[84]在面向物体识别进行特征提取时，基于可变形卷积[59]不同时刻的脉冲特征进行了对齐；SNNTracker[85]在进行多目标跟踪时，利用神经元模型提取特征，动态适配脉冲流的特征。这些思路均体现了脉冲影像重建工作对于下游视觉任务的指导作用。

5　融合脉冲相机的多传感器成像算法

脉冲相机基于“积分－发放”机制实现了超高速场景下的连续光强记录，其通过脉冲频率差异精准反映光强变化，本身具备高动态成像能力，但其

性能受限于硬件设计中数据带宽与存储效率的权衡，脉冲相机的空间分辨率通常相对较低，难以满足高精度视觉任务对于静态纹理细节的需求，传统帧式相机则与之形成互补：虽其动态范围受限于传感器物理特性，却常具备高空间分辨率，可提供丰富的静态纹理；事件相机则同样拥有高动态范围与低延迟特性，仅记录光强变化的“事件”信号，无法直接获取静态区域的完整纹理，难以独立支撑需要全局场景信息的成像任务。

这些传感器在性能上的互补特性，推动研究者探索融合神经形态相机的多传感器成像系统[86-88]。如图 9 所示，基于分光棱镜的双相机成像系统可实现不同相机的优势互补。在脉冲相机领域，通过整合脉冲相机的高时间分辨率与高动态优势，与传统相机的高空间分辨率、事件相机的高动态范围等特性，在弥补脉冲相机空间分辨率不足的同时，解决传统相机动态范围有限、事件相机纹理缺失等问题，最终实现复杂场景下“高时间分辨率－高空间分辨率－高动态范围”的协同优化。

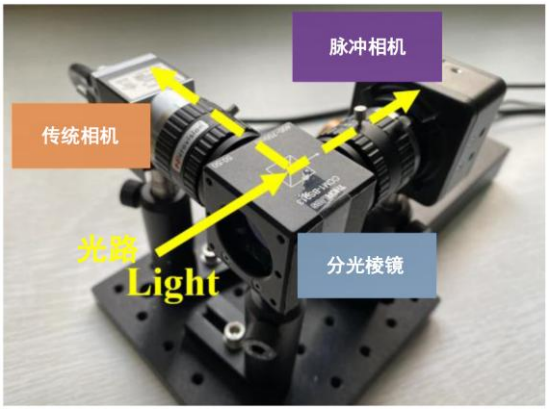

图 9　融合双相机的成像系统[89]

本节将围绕融合脉冲相机的多传感器成像算法展开，重点梳理三个核心研究方向。首先聚焦最基础的面向高帧率成像的算法，解决高帧率下“时间分辨率与空间分辨率协同”的基础需求；在此之上，进一步拓展至面向高动态成像的融合方案与面向去模糊成像的融合方案，分别应对复杂光照与动态模糊问题。下文将依次介绍三类方案的典型算法，展示多传感器融合如何为脉冲成像技术的实用化突破提供关键支撑。

5.1　面向高帧率成像的算法

在融合脉冲相机的多传感器高动态成像研究中，核心思路围绕“优势互补”：脉冲相机凭“积分－发放”机制具备高动态成像能力，可精准捕捉光强变化，但空间分辨率不足、静态纹理呈现有限；传统帧式相机动态范围受物理特性约束，却能提供高分辨率静态纹理。面向高帧率成像的算法旨在融合两者优势，实现高帧率成像。

Xia 等人[89]提出的 SVFI 方案，通过融合脉冲相机与传统帧式相机的多模态信息，应对了高速动态场景下视频插帧中因运动估计失效导致的时序信息断层与遮挡伪影问题，最终实现高帧率成像。该方案的核心是利用脉冲相机对相邻两帧 RGB 影像间光强变化的连续记录能力，弥补传统相机固定帧率下时域信息缺失的短板，为高帧率插帧提供充足动态时序支撑。

Geng 等人[90]提出超高速高光谱成像系统，融合压缩感知与神经形态采样，突破传统高光谱成像“静态局限”与高速成像“无光谱信息”的矛盾，实现高帧率下高光谱信息精准捕获。该系统采用双分支硬件架构：压缩采样分支以编码孔径快照光谱成像仪捕获中等帧率高光谱测量帧，完整保留场景光谱维度信息；神经形态分支通过脉冲相机输出高帧率密集灰度脉冲流，精准记录高速动态场景的连续时空细节，两类数据形成“光谱信息+高帧率动态”的双模态互补，为高帧率高光谱成像提供基础。

重建框架核心是脉冲光谱先验网络（SSPN）与双模态即插即用（PnP）迭代优化，聚焦保障高帧率下高光谱重建质量。其中，SSPN 包含脉冲超分辨率重建子网络（R－子网络）和脉冲引导高光谱先验去噪子网络（D－子网络）：R－子网络采用混合 SNN-ANN 架构，从脉冲流中同步提取静态空间特征与动态时序特征，生成高分辨率灰度帧，为高帧率时序对齐提供清晰参考；D－子网络通过模态特征对齐、跨模态先验迁移与低秩先验增强，将脉冲流中的高帧率时序信息转化为高光谱重建的强先验，修正噪声干扰与光谱失真，确保高帧率下光谱细节不丢失。PnP 迭代优化交替执行数据保真项与先验正则项，前者确保重建结果与硬件观测数据的一致性，后者依托 D－子网络注入脉冲引导先验，最终实现兼具高光谱分辨率与高时间分辨率的高帧率高光谱影像重建。

本节的高帧率成像融合方法主要回应了第 3.4 节中“单一传感器难以同时兼顾高时间分辨率与高空间分辨率”的挑战，通过由脉冲相机提供连续时间线索、传统相机提供高分辨率纹理，通过跨模态数据融合实现高帧率视频重建。

5.2　面向高动态成像的算法

在高帧率成像基础上，面向高动态成像的融合

方案进一步聚焦“复杂光照下的细节保留”，通过多传感器信息互补突破单一设备的动态范围局限。脉冲相机凭“积分－发放”机制具备了一定高动态能力，可捕捉光强细微变化，但强光区域易因脉冲积分过饱和而丢失细节，弱光区域则因脉冲稀疏而导致光强估计误差较大；传统帧式相机动态范围受限于传感器物理阈值，强光过曝、弱光欠曝问题突出；事件相机虽能记录极宽光强范围的变化，但无法直接获取静态区域纹理。

基于上述互补特性，面向高动态成像的融合方案形成两类路线：其一为脉冲相机与传统帧式相机融合，以脉冲相机的高动态光强记录为基础，通过传统相机补充高分辨率纹理细节；其二为脉冲相机与事件相机融合，借助事件相机的高动态特性进一步突破光强记录极限，同时依靠脉冲相机提供静态纹理信息，弥补事件相机全局场景信息不足的短板。两类方案均以多模态信息协同为核心，最终有效改善强光过曝、弱光欠曝问题，下文将具体梳理两类方案的技术原理与典型算法。

Han 等人[30]提出神经形态相机引导的高动态范围成像方案，核心聚焦单张高动态范围影像重建。其将传统相机输出的低动态范围影像经逆相机响应函数转换至线性域并变为 YUV 色彩空间，利用 Y 通道与神经形态相机（如脉冲相机、事件相机）输出的高动态强度图融合。为处理强度图与低动态范围影像分辨率差异，采用反卷积层上采样强度图，在亮度域通过设计线性斜坡加权函数和注意力掩码模块，优先保留传统相机有效纹理，同时借助强度图补充过曝、欠曝区域细节，最后经基于残差块的自编码器进行色度补偿，输出高动态且纹理完整的彩色高动态影像。Han 等人[91]后续将该方法扩展至高动态范围视频重建，并增加循环模块保障视频帧间时间一致性。

Zhu 等人[92]针对脉冲相机与事件相机的性能互补特性，提出 NeuSpike-Net 实现高动态、高速场景下的高质量成像。该方法以人类视网膜“中央凹－外周”协同机制为灵感，模拟中央凹的脉冲相机负责提供完整纹理信息、模拟外周的事件相机负责拓展动态范围，通过三部分设计实现两类神经形态数据有效融合：首先，设计自适应数据表征方式，为脉冲流构建“脉冲间隔 + 脉冲平面”输入形式以保留时域光强与纹理关联，为事件流基于积分－发放模型设计累积表征以转换异步事件为二维动态特征；其次，搭建双路径 U 型网络架构，分别通过运动路径提取事件流的高动态与运动特征、通过纹理路径提取脉冲流的纹理特征，并借助注意力驱动的特征融合模块在编码过程中完成通道与空间维度的协同，实现高动态细节与完整纹理的恢复。

Chang 等人[93]提出基于脉冲－RGB 混合相机的 1000 FPS 高动态范围视频成像方案。其路线为脉冲相机提供运动细节、传统相机提供纹理色彩与长短曝光信息。传统相机的长短时长交替曝光虽能覆盖部分动态范围，但长曝光易模糊、短曝光易欠曝，难以兼顾高光与低光细节；脉冲相机可连续记录光强变化，捕捉长曝光过程中运动信息。

该方案通过三阶段框架实现融合：一是脉冲预处理，通过光流对齐，结合帧融合与像素重排提升空间分辨率，抑制噪声的同时保留高动态细节；二是 RGB 预处理，线性映射统一亮度后，用脉冲引导去模糊，以脉冲中的运动信息补偿中 / 长曝光帧中的运动模糊；三是融合生成高动态范围视频，通过 CNN-RNN 模块整合脉冲高动态特征与 RGB 色彩纹理，通过 ConvLSTM 维持帧间一致性，最终输出 1000 FPS 的高动态视频，实现了高帧率与高动态范围的融合。

面向高动态成像的融合方案主要对应第 3.3、3.4 节提出的“单机动态范围有限、强光过曝与弱光欠曝并存”的问题，利用脉冲 / 事件相机记录宽动态光强、传统相机补充细节纹理，在多模态联合重建中恢复高动态范围影像。

5.3 面向去模糊成像的算法

在高速动态场景成像中，传统帧式相机受曝光机制限制，易因相机运动或目标高速移动产生运动模糊，同时大光圈拍摄导致的浅景深会引发局部区域离焦模糊，两类模糊共同制约成像清晰度。受数据带宽限制，脉冲相机的空间分辨率通常有限，但能记录细致运动过程。

基于上述互补特性，融合脉冲相机的多传感器去模糊方案形成两类路线：一类是面向运动模糊的融合方案，依托脉冲相机高时间分辨率脉冲流提取精准运动信息建模动态过程，结合传统相机高分辨率纹理实现高速动态场景清晰成像；另一类是面向离焦模糊的融合方案，利用脉冲相机在焦平面扫掠中记录的连续光强变化，辅助构建多深度聚焦信息，弥补单张离焦影像的深度细节缺失。下文将介绍两类方案的典型算法，展示脉冲相机与其他传感器协同突破单一设备性能局限的过程。

Chen 等人[94]提出的 SpkDeblurNet，实现了脉

冲相机与传统帧式相机融合的运动去模糊模型。如图 10 所示，其针对传统帧式相机在高速场景下因曝光窗口产生的运动模糊，及脉冲相机空间分辨率低、无色彩信息的短板，通过双分支架构与专用融合模块实现多模态协同去模糊。其采用双分支 Transformer 架构：RGB 分支通过残差 Swin Transformer 块（RSTB）提取高分辨率色彩与纹理特征，脉冲分支经 RSTB 挖掘脉冲流时空信息；为高效跨模态融合，设计内容感知运动幅度注意力模块，从 RGB 初始去模糊结果中自适应筛选清晰区域纹理，生成运动掩码引导脉冲分支重建高分辨率灰度图；同时引入转置交叉注意力融合模块，沿通道维度实现两类特征双向对齐与加权，避免传统空间注意力的计算冗余。该方法通过联合训练去模糊与脉冲重建任务，消除高速运动模糊的同时兼顾纹理完整性与色彩保真度，形成了“双向特征互补－高效融合－联合优化”的去模糊范式。

此前双相机融合方案多依赖两相机严格对齐，常需分光棱镜组成的光学系统，给实际部署带来困难。Zhang 等人[95]提出的 UaSDN 算法，突破上述相机对齐的严格假设，针对真实场景中两类传感器因光学校准复杂、设备振动位移、时域同步困难导致的时空错位问题，通过三阶段渐进式架构与专用对齐模块，实现“无需预校准即可高效融合脉冲与 RGB 信息”的去模糊优化。

UaSDN 的设计围绕“渐进优化+自适应对齐”开展。第一阶段采用 NAFNet[96]作为基础去模糊网络，对模糊 RGB 影像进行初步去模糊与纹理增强，输出基础清晰影像。第二阶段针对时空错位，设计双向可变形模态对齐模块，先通过浅层卷积从脉冲流中估计粗粒度光强图，再通过多尺度可变形卷积实现“RGB 特征向脉冲特征对齐→跨模态特征融合→融合特征向 RGB 特征回对齐”的双向优化，动态学习两类数据的空间偏移与时域偏差。第三阶段构建光流引导的多尺度密集融合网络，计算脉冲粗光强图与对齐后影像的光流并完成光流映射，结合注意力密集块实现多尺度特征融合与细节补全，最终输出高质量去模糊影像。该方法突破对齐约束对硬件部署的限制，为高速场景下脉冲－RGB 混合去模糊系统实用化提供关键技术。

上述脉冲与传统相机融合方案中，深度神经网络常依赖监督学习训练，需大规模“模糊－清晰”配对数据集，而这类数据收集困难：实拍数据拍摄难，合成数据与真实场景存在差异。为缓解该问题，Chen 等人[57]提出基于无监督学习的脉冲与传统相机融合去模糊成像方案。该无监督框架通过双分支特征提取与多约束协同优化实现去模糊：RGB 分支采用轻量级残差网络对模糊影像进行初步去模糊；脉冲分支通过时序卷积提取脉冲流中高时间分辨率运动特征，转化为运动掩码与光强参考图。

该方法包含三类无监督损失：一是跨模态一致性损失，约束 RGB 分支输出的亮度通道与脉冲分支构建的光强参考图在梯度与全局光强分布上的一致性；二是自监督去模糊损失，对候选清晰影像添加随机运动模糊核生成“伪模糊影像”，计算其与输入模糊 RGB 影像的差异，优化去模糊网络；三是模态对齐损失，利用可变形卷积动态实现脉冲与 RGB 数据的时空对齐，通过特征互信息最大化确保跨模态特征匹配。通过上述损失函数，实现基于无监督学习的脉冲－传统混合相机去模糊成像，为其在缺乏标注数据的真实场景中应用提供支撑。

在运动强度较大的动态场景中，传统全聚焦成像面临场景适用性与成像质量双重挑战：一方面，多帧拍摄构建焦栈融合全聚焦影像的方式，要求场景静态且需精心设计聚焦距离，难应对动态场景；另一方面，单帧散焦去模糊因散焦核受光圈形状、场景深度影响呈空间变化，存在高度病态性，易在弱纹理、强散焦区域产生振铃伪影。

神经形态相机因超高时间分辨率，可进行细致焦平面扫掠记录，为离焦模糊补充多深度聚焦信息。Teng 等人[97]提出针对高速动态场景离焦模糊的脉冲－传统帧式相机融合方法。其利用脉冲相机单次曝光内焦平面连续移动的特性，将脉冲流转换为多深度聚焦置信度图，标记不同区域聚焦程度；通过深度感知对齐模块匹配脉冲与 RGB 影像空间位置，再以聚焦置信度为指导，对 RGB 离焦区域分层增强，保留清晰区域纹理的同时，依托脉冲重建离焦细节。该方案无需多帧对焦影像，仅通过单张 RGB 与同步的脉冲流，即可实现高速场景下离焦模糊消除。

去模糊成像相关方法则针对第 3.2、3.4 节中的“高速运动下传统帧式成像严重模糊、脉冲与 RGB 难以精确对齐”等挑战，借助脉冲流提供的高时间分辨率约束和专用对齐－融合模块，在消除运动模糊的同时尽量保留细节与色彩信息。

综上所述，融合成像的核心在于借助脉冲相机高时间分辨率、高动态特性，与传统帧式相机高空间分辨率、事件相机高动态优势互补，以突破单一

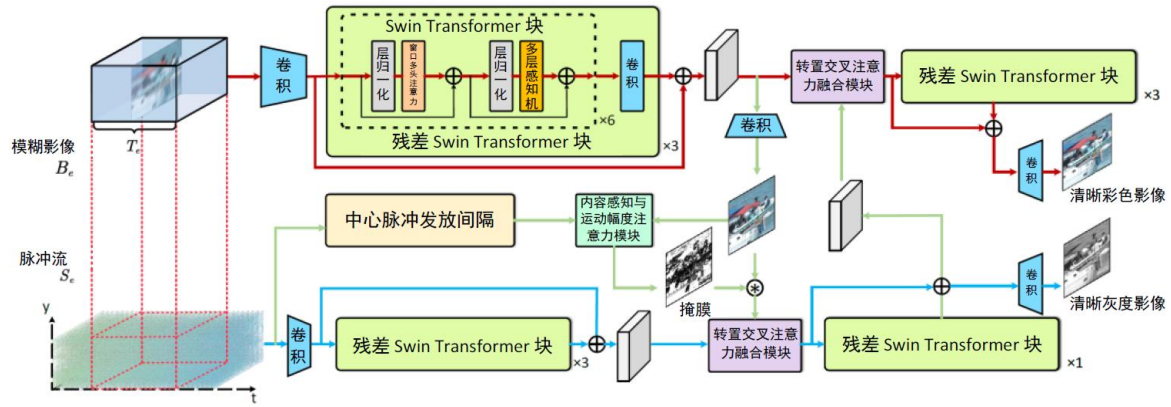

图 10　Chen 等人[94]提出的脉冲-传统相机融合去模糊算法 SpkDeblurNet

传感器局限。在高动态成像方面，现有方法主要通过脉冲提供高动态信息、脉冲为长短曝光提供运动信息、脉冲与事件融合这三类路线，针对性优化复杂光照与高速场景下的成像效果，有效改善过曝与欠曝问题。在去模糊方向，针对运动与离焦模糊，脉冲相机提供关键支撑：以脉冲流提取运动信息建模动态过程，辅助传统相机重建模糊细节；靠焦平面扫掠记录构建多深度信息，弥补离焦影像深度缺失。这类方案挖掘脉冲信号价值，为高速动态场景成像提供技术支撑。

6 结语与展望

6.1 结语

本文综述了超高速脉冲相机的成像方法。脉冲相机旨在解决传统帧式相机在高速场景下的“短曝光噪声与长曝光模糊”问题。脉冲相机基于“积分-发放”原理，每个像元独立累积光子并异步发放脉冲，这使其区别于传统相机的全局同步曝光与事件相机的光强变化采样方式。凭借上述特性，脉冲相机能够同时保留静态纹理与动态运动信息。脉冲相机成像算法经历了从时序统计到深度建模的演变，解决了噪声抑制与运动对齐问题，拓展了其应用范围至超分辨率、彩色、高动态、三维成像等领域。

多传感器融合算法通过“优势互补”突破单设备的局限，旨在解决脉冲相机空间分辨率不足、传统相机动态范围和帧率受限的问题。高帧率方案融合脉冲流与传统相机纹理，缓解了高帧率成像与高分辨率之间的矛盾；高动态与去模糊方案则结合脉冲相机与传统相机、事件相机的优势，实现了动态范围拓展与模糊消除，最终实现了“高时间分辨率－高空间分辨率－高动态范围”的协同优化。

6.2 脉冲成像的未来研究展望

脉冲成像技术在突破传统帧式相机局限、拓展高速动态场景成像能力上已展现显著潜力。然而，面向更复杂的实际需求与技术壁垒，仍需从数据支撑、技术维度、智能融合、硬件适配等多层面深化探索，结合现有方向，可进一步拓展以下研究方向。

大型真实数据集的构建。当前脉冲成像算法的训练与验证多依赖合成数据或小规模实拍数据，合成数据与真实脉冲流的差异限制了算法的泛化能力[98-99]。未来应推动多场景、多模态、标准化脉冲数据集的建设，设计涵盖高速运动、极端光照、复杂纹理等的实拍方案，结合摩尔纹[100]、水下成像[101]等特殊场景，为脉冲成像算法提供数据基础。

脉冲光场成像技术的构建。光场是描述空间中所有光线传播状态的物理概念[102-103]，包含光线的空间位置、传播方向及光强等信息。脉冲光场成像技术可通过脉冲相机与多视角成像系统的集成，利用脉冲流记录不同空间位置和视角的光强时序变化，实现场景的三维信息采集。在算法上，突破传统光场成像的静态场景限制，构建基于脉冲流的动态建模方法，实现动态视角渲染与三维结构重建。

语义信息驱动的脉冲成像。当前脉冲成像多停留在信号解析层面，缺乏对场景语义的挖掘[104]。未来可推动语义信息与脉冲成像融合，构建协同框架，将场景语义融入重建过程中，实现对关键语义区域的优先保留，并抑制背景噪声。

脉冲成像硬件与算法的协同优化。未来可开发可调阈值的自适应脉冲传感器，并设计轻量化模型以适配硬件特性[105-106]，推动脉冲成像系统向“高时效、低功耗、小型化”发展，满足移动端、嵌入式设备的需求。

跨领域脉冲成像应用的拓展与定制化。脉冲成

像的应用多集中在通用场景，未来可向垂直领域深入，如工业检测中的零件缺陷检测、医疗诊断中的生物组织动态成像、航空航天中的飞行器动态检测等，进一步拓展其应用场景。

参 考 文 献

[1] Huang T, Zheng Y, Yu Z, et al. 1000x faster camera and machine vision with ordinary devices. Engineering, 2023, 25: 110-119.

[2] Huang T, Yu Z, Li Y, et al. Advances in spike vision. Journal of Image and Graphics, 2022, 27(6): 1823-1839.
(黄铁军, 余肇飞，李源，等. 脉冲视觉研究进展. 中国图象图形学报, 2022, 27(6) : 1823-1839.)

[3] Li J, Tian Y. Recent advances in neuromorphic vision sensors: A survey . Chinese Journal of Computers, 2021, 44(6): 1258-1286.
(李佳宁, 田永鸿. 神经形态视觉传感器的研究进展及应用综述. 计算机学报, 2021, 44(6): 1258-1286.)

[4] Zhu L, Tian Y. Review of visual reconstruction methods of retina-like vision sensors. Scientia Sinica Informationis, 2023, 53(3): 417-436.
(朱林, 田永鸿. 仿视网膜传感器视觉重建算法研究综述. 中国科学：信息科学, 2023, 53(3): 417-436.)

[5] Yu L, Shi B, Wang W, et al. Neuromorphic-enabled visual enhancement: principles, methods and recent advances. Journal of Image and Graphics, 2025, 30(6): 1593-1615.
(余磊, 施柏鑫，王威，等. 类脑赋能视觉增强：原理、方法与前沿进展. 中国图象图形学报, 2025, 30(6): 1593-1615.)

[6] Dong S, Huang T, Tian Y. Spike camera and its coding methods// Proceedings of the Data Compression Conference. Snowbird, USA, 2017: 437-437.

[7] Lichtsteiner P, Posch C, Delbruck T. A 128times128 120 db 15mu s latency asynchronous temporal contrast vision sensor. IEEE Journal of Solid-State Circuits, 2008, 43(2): 566-576.

[8] Brandli C, Berner R, Yang M, et al. A 240 180 130 db 3 μs latency global shutter spatiotemporal vision sensor. IEEE Journal of Solid-State Circuits, 2014, 49(10): 2333-2341.

[9] Zheng Y, Zhang J, Zhao R, et al. Spikecv: open a continuous computer vision era. 2023. https://arxiv.org/abs/2303.11684.

[10] Hu L, Zhao R, Ding Z, et al. Optical flow estimation for spiking camera//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. New Orleans, USA, 2022: 17844-17853.

[11] Zhao R, Xiong R, Zhao J, et al. Learning optical flow from continuous spike streams. Advances in Neural Information Processing Systems, 2022, 35: 7905-7920.

[12] Xia L, Ding Z, Zhao R, et al. Unsupervised optical flow estimation with dynamic timing representation for spike camera//Advances in Neural Information Processing Systems. New Orleans,USA , 2023: 1-13.

[13] Zhao R, Xiong R, Zhang J, et al. Optical flow for spike camera with hierarchical spatial-temporal spike fusion//Proceedings of the AAAI Conference on Artificial Intelligence: Vol. 38. Vancouver, Canada, 2024: 7496-7504.

[14] Li J, Wang X, Zhu L, et al. Retinomorphic object detection in asynchronous visual streams//Proceedings of the AAAI Conference on Artificial Intelligence. Virtual, 2022: 1332-1340.

[15] Zheng Y, Yu Z, Wang S, et al. Spike-based motion estimation for object tracking through bio-inspired unsupervised learning. IEEE Transactions on Image Processing, 2022, 32: 335-349.

[16] Wang Y, Li J, Zhu L, et al. Learning stereo depth estimation with bio-inspired spike cameras//Proceedings of the IEEE International Conference on Multimedia and Expo. Taipei, China, 2022: 1-6.

[17] Zhang J, Tang L, Yu Z, et al. Spike Transformer: Monocular depth estimation for spiking camera//Proceedings of the European Conference on Computer Vision. Tel Aviv, Israel ,2022: 34-52.

[18] Zhu L, Dong S, Huang T, et al. A retina-inspired sampling method for visual texture reconstruction//Proceedings of the IEEE International Conference on Multimedia and Expo. Shanghai， China, 2019: 1432-1437.

[19] Zhao J, Xiong R, Xie J, et al. Reconstructing clear image for high-speed motion scene with a retina-inspired spike camera. IEEE Transactions on Computational Imaging, 2021, 8: 12-27.

[20] Zhao J, Xiong R, Liu H, et al. Spk2ImgNet: Learning to reconstruct dynamic scene from continuous spike stream//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville，USA, 2021: 11996-12005.

[21] MAASS W. Networks of spiking neurons: the third generation of neural network models. Neural Networks, 1997, 10(9): 1659-1671.

[22] Li W, Zhao X L, Ma Z, et al. Motion-decoupled spiking transformer for audio-visual zero-shot learning//Proceedings of the ACM International Conference on Multimedia. New York，USA, 2023: 3994-4002.

[23] Li W, Wang P, Xiong R, et al. Spiking tucker fusion transformer for audio-visual zero-shot learning. IEEE Transactions on Image Processing, 2024, 33: 4840-4852.

[24] Zhu L, Dong S, Li J, et al. Retina-like visual image reconstruction via spiking neural model//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA, 2020: 1438-1446.

[25] Zhao J Xie J, Xiong R, et al. Super resolve dynamic scene from continuous spike streams//Proceedings of the IEEE/CVF International Conference on Computer Vision. Montreal, USA, 2021: 2533-2542.

[26] Dong Y, Zhao J, Xiong R, et al. 3D residual interpolation for spike camera demosaicing//Proceedings of the IEEE International Conference on Image Processing. Bordeaux, France, 2022: 1461-1465.

[27] Xiong B, Wang R, Zhou Y, et al. High-speed phase transition imaging using spike streams//Adaptive Optics: Methods, Analysis and Applications. Toulouse, France, 2024: JD6A-2.

[28] Mildenhall B, Srinivasan P P, Tancik M, et al. NeRF: Representing

scenes as neural radiance fields for view synthesis//European Conference on Computer Vision. Glasgow, UK, 2020: 405-421.

[29] Kerbl B, Kopanas G, LEIMKÜHLER T, et al. 3D Gaussian splatting for real-time radiance field rendering. ACM Transactions on Graphics, 2023, 42(4): 1-14.

[30] Han J, Zhou C, Duan P, et al. Neuromorphic camera guided high dynamic range imaging//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA, 2020: 1730-1739.

[31] Huang T. Spiking continuous photographing principle and demonstration on ultrahigh speed and high dynamic imaging. Acta Electronica Sinica, 2022, 50(12): 2919-2927.
(黄铁军. 脉冲连续摄影原理与超高速高动态成像验证. 电子学报, 2022, 50(12): 2919-2927.)

[32] Zhang J, Zheng Y, Yu Z, et al. Spike-based vision for autonomous driving scenarios. Strategic Study of CAE, 2024, 26(1): 160-177.
(张济远，郑雅菁，余肇飞，黄铁军. 面向自动驾驶场景的脉冲视觉研究. 中国工程科学, 2024, 26(1): 160-177.)

[33] Zhao R, Xiong R, Zhao J, et al. Boosting spike camera image reconstruction from a perspective of dealing with spike fluctuations//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA, 2024: 24955-24965.

[34] Horn B K, Schunck B G. Determining optical flow. Artificial intelligence, 1981, 17(1-3): 185-203.

[35] Zhao J, Xiong R, Huang T. High-speed motion scene reconstruction for spike camera via motion aligned filtering//Proceedings of the IEEE International Symposium on Circuits and Systems. Seville, Spain, 2020: 1-5.

[36] Brox T, Bruhn A, Papenberg N, et al. High accuracy optical flow estimation based on a theory for warping//European Conference on Computer Vision. Prague, Czech Republic, 2004: 25-36.

[37] Zhang Y, Xiong R, Huang T. Spike signal reconstruction based on inter-spike similarity//IEEE International Conference on Visual Communications and Image Processing. Suzhou, China, 2022: 1-5.

[38] Wang Y, Xiong R, Zhao J, et al. Reconstruct dynamic scene for spike camera based on 3d space time similarity//Proceedings of the IEEE International Conference on Image Processing. Abu Dhabi, United Arab Emirates, 2024: 1595-1601.

[39] Dong Y, Zhao J, Xiong R, et al. High-speed scene reconstruction from low-light spike streams//IEEE International Conference on Visual Communications and Image Processing. Suzhou, China, 2022: 1-5.

[40] Zhu L, Dong S, Li J, et al. Ultra-high temporal resolution visual reconstruction from a fovea-like spike camera via spiking neuron model. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 45(1): 1233-1249.

[41] Bi G Q, Poo M M. Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. Journal of Neuroscience, 1998, 18(24): 10464-10472.

[42] Song S, Miller K D, Abbott L F. Competitive hebbian learning through spike-timing-dependent synaptic plasticity. Nature Neuroscience, 2000, 3(9): 919-926.

[43] Zheng Y, Zheng L, Yu Z, et al. High-speed image reconstruction through short-term plasticity for spiking cameras//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA, 2021: 6358-6367.

[44] Tsodyks M V, Markram H. The neural code between neocortical pyramidal neurons depends on neurotransmitter release probability. Proceedings of the National Academy of Sciences, 1997, 94(2): 719-723.

[45] Tsodyks M, Pawelzik K, Markram H. Neural networks with dynamic synapses. Neural computation, 1998, 10(4): 821-835.

[46] Zheng Y, Zheng L, Yu Z, et al. Capture the moment: High-speed imaging with spiking cameras through short-term plasticity. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(7): 8127-8142.

[47] Chen K, Chen S, Zhang J, et al. SpikeReveal: Unlocking temporal sequences from real blurry inputs with spike streams. Advances in Neural Information Processing Systems, 2024, 37: 1-14.

[48] Wang X, Chan K C, Yu K, et al. EDVR: Video restoration with enhanced deformable convolutional networks//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Long Beach, USA, 2019: 1-10.

[49] Dai J, Qi H, Xiong Y, et al. Deformable convolutional networks//Proceedings of the IEEE International Conference on Computer Vision. Venice, Italy, 2017: 764-773.

[50] Zhu X, Hu H, Lin S, et al. Deformable convnets v2: More deformable, better results//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA, 2019: 9308-9316.

[51] Chen S, Duan C, Yu Z, et al, Self-supervised mutual learning for dynamic scene reconstruction of spiking camera//Proceedings of the International Joint Conference on Artificial Intelligence. Vienna, Australia, 2022: 2859-2866.

[52] Chen S, Yu Z, Huang T. Self-supervised joint dynamic scene reconstruction and optical flow estimation for spiking camera//Proceedings of the AAAI Conference on Artificial Intelligence. Washington DC, USA, 2023: 350-358.

[53] Zhang J, Jia S, Yu Z, et al. Learning temporal-ordered representation for spike streams based on discrete wavelet transforms//Proceedings of the AAAI Conference on Artificial Intelligence. Washington DC, USA, 2023: 137-147.

[54] Zhu L, Zheng Y, Geng M, et al. Recurrent spike-based image restoration under general illumination//Proceedings of the ACM International Conference on Multimedia. Ottawa, Canada, 2023: 8251-8260.

[55] Hu L, Ding Z, Liu M, et al. Learning to robustly reconstruct dynamic scenes from low-light spike streams// Proceedings of the European Conference on Computer Vision. Milano, Italy, 2024: 88-105.

[56] Zhao R, Xiong R, Zhang J, et al. Spike camera image reconstruction using deep spiking neural networks. IEEE Transactions on Circuits and Systems for Video Technology, 2024, 34(6): 5207-5212.

[57] Fan B, Yin J, Dai Y, et al. Spatio-temporal interactive learning for efficient image reconstruction of spiking cameras//Proceedings of the Annual Conference on Neural Information Processing Systems. Vancouver, Canada, 2024: 1-14.

[58] Yang C, Li G, Wang S, et al. SpikeODE: Image reconstruction for spike camera with neural ordinary differential equation. IEEE Transactions on Circuits and Systems for Video Technology, 2024, 34(11): 11142-11155.

[59] Yin J, Fan B, Xu C, et al. Spk2ImgMamba: Spiking camera image reconstruction with multi-scale state space models//2025 IEEE/CVF Winter Conference on Applications of Computer Vision. 2025: 9342-9352.

[60] Chen K, Zheng Y, Huang T, et al. Rethinking high-speed image reconstruction framework with spike camera//Proceedings of the AAAI Conference on Artificial Intelligence: Vol. 39. Philadelphia, USA, 2025: 2097-2104.

[61] Lehtinen J, Munkberg J, Hasselgren J, et al. Noise2Noise: Learning image restoration without clean data// Proceedings of the International Conference on Machine Learning. Stockholm, Sweden ,2018: 4620-4631.

[62] Ding X, Zhang X, Ma N, et al. RepVGG: Making vgg-style convnets great again//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA, 2021: 13733-13742.

[63] GU A, GOEL K, RE C. Efficiently modeling long sequences with structured state spaces//International Conference on Learning Representations. Virtual, 2022: 1-15.

[64] Radford A, Kim J W, Hallacy C, et al. Learning transferable visual models from natural language supervision//International Conference on Machine Learning. Virtual, 2021: 8748-8763.

[65] Xiang X, Zhu L, Li J, et al. Learning super-resolution reconstruction for high temporal resolution spike stream. IEEE Transactions on Circuits and Systems for Video Technology, 2021, 33(1): 16-29.

[66] Zhao J, Xiong R, Zhang J, et al. Learning to super-resolve dynamic scenes for neuromorphic spike camera//Proceedings of the AAAI Conference on Artificial Intelligence: Vol. 37. Washington , USA, 2023: 3579-3587.

[67] Wang Y, Zhang Y, Xiong R, et al. Spk2srimgnet: Super-resolve dynamic scene from spike stream via motion aligned collaborative filtering//Proceedings of the Computer Vision and Pattern Recognition Conference. Nashville, USA, 2025: 11416-11426.

[68] Dong Y, Xiong R, Zhao J, et al. Learning a deep demosaicing network for spike camera with color filter array. IEEE Transactions on Image Processing, 2024, 33: 3634-3647.

[69] Teed Z, Deng J. Raft: Recurrent all-pairs field transforms for optical flow// Proceedings of the European Conference on Computer Vision. Virtual, 2020: 402-419.

[70] Dong Y, Xiong R, Zhao J, et al. Joint demosaicing and denoising for spike camera//Proceedings of the AAAI Conference on Artificial Intelligence: Vol. 38. Vancouver, Canada, 2024: 1582-1590.

[71] Dong Y, Xiong R, Zhang J, et al. Super-resolution reconstruction from bayer-pattern spike streams//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA, 2024: 24871-24880.

[72] Yang S, Huang Z, Chang Y, et al. Real-data-driven 2000 fps color video from mosaicked chromatic spikes// Proceedings of the European Conference on Computer Vision. Springer, 2024: 305-321.

[73] Dong Y, Xiong R, Fan X, et al. Self-supervised learning for color spike camera reconstruction//Proceedings of the Computer Vision and Pattern Recognition Conference. Nashville, USA, 2025: 6231-6240.

[74] Dong Y, Xiong R, Fan X, et al. Dynamic scene reconstruction for color spike camera via zero-shot learning. IEEE Transactions on Computational Imaging, 2025, 11: 120-141.

[75] Chang Y, Xiaokaiti Y, Liu Y, et al. Towards HDR and HFR video from rolling-mixed-bit spikings//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA, 2024: 25117-25127.

[76] Zhu Z, Xiong R, Zhao J, et al. High dynamic range imaging for dynamic scenes based on multi-level spike camera. IEEE Transactions on Circuits and Systems for Video Technology, 2025, 35(6): 5394-5406.

[77] Gu A, Johnson I, Goel K, et al. Combining recurrent, convolutional, and continuous-time models with linear state space layers//Advances in Neural Information Processing Systems. Virtual, 2021: 572-585.

[78] Xiong B, Su C, Lin Z, et al. Real-time parameter evaluation of high-speed microfluidic droplets using continuous spike streams//Proceedings of the ACM International Conference on Multimedia. New York, USA, 2024: 6833-6841.

[79] Zhang J, Chen S, Zheng Y, et al. Transient glimpses: unveiling occluded backgrounds through the spike camera//Proceedings of the AAAI Conference on Artificial Intelligence: Vol. 38. Vancouver, Canada, 2024: 637-645.

[80] Zhu L, Jia K, Zhao Y, et al. SpikeNeRF: Learning neural radiance fields from continuous spike stream//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA, 2024: 6285-6295.

[81] Zhang J, Chen K, Chen S, et al. SpikeGS: 3d gaussian splatting from spike streams with high-speed camera motion//Proceedings of the ACM International Conference on Multimedia. Melbourn, Australia, 2024: 9194-9203.

[82] Yu J, Peng X, Lu Z, et al. SpikeGS: Learning 3d gaussian fields from continuous spike stream//Proceedings of the Asian Conference on Computer Vision. Hanoi, Vietnam, 2024: 4280-4298.

[83] Guo Y, Hu L, Bai Y, et al. SpikeGS: Reconstruct 3d scene via fast-moving bio-inspired sensors//Proceedings of the AAAI Conference on Artificial Intelligence. Philadelphia, USA, 2025:

3293-3301.

[84] Zhao J, Zhang S, Yu Z, et al. Spireco: Fast and efficient recognition of high-speed moving objects with spike camera. IEEE Transactions on Circuits and Systems for Video Technology, 2023, 34(7): 5856-5867.

[85] Zheng Y, Li C, Zhang J, et al. SNNTracker: Online high-speed multi-object tracking with spike camera. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2026, 48(1): 624 - 638.

[86] Xia L, Zhao J, Xiong R, et al. SVFI: spiking-based video frame interpolation for high-speed motion//Proceedings of the AAAI Conference on Artificial Intelligence: Vol. 37. Washington DC, USA, 2023: 2910-2918.

[87] Wang Z W, Duan P, Cossairt O, et al. Joint filtering of intensity images and neuromorphic events for high-resolution noise-robust imaging//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA, 2020: 1609-1619.

[88] Xiao Z, Kai D, Zhang Y, et al. Asymmetric event-guided video super-resolution//Proceedings of the ACM International Conference on Multimedia. Melbourne, Australia, 2024: 2409-2418.

[89] Xiao Z, Wang X. Event-based video super-resolution via state space models//Proceedings of the Computer Vision and Pattern Recognition Conference. Nashville, USA, 2025: 12564-12574.

[90] Geng M, Wang L, Zhu L, et al. Towards ultra high-speed hyperspectral imaging by integrating compressive and neuromorphic sampling. International Journal of Computer Vision, 2025, 133(4): 1587-1610.

[91] Han J, Yang Y, Duan P, et al. Hybrid high dynamic range imaging fusing neuromorphic and conventional images. IEEE Transactions on pattern analysis and machine intelligence, 2023, 45(7): 8553-8565.

[92] Zhu L, Li J, Wang X, et al. NeuSpike-Net: High speed video reconstruction via bio-inspired neuromorphic cameras//Proceedings of the IEEE/CVF international conference on computer vision. Montreal, USA , 2021: 2400-2409.

[93] Chang Y, Zhou C, Hong Y, et al. 1000 FPS HDR video with a Spike-RGB hybrid camera//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada, 2023: 22180-22190.

[94] Chen S, Zhang J, Zheng Y, et al. Enhancing motion deblurring in high-speed scenes with spike streams//Advances in Neural Information Processing Systems: Vol. 36. New Orleans, USA, 2023: 1-14.

[95] Zhang J, Chen S, Zheng Y, et al. Spike-guided motion deblurring with unknown modal spatiotemporal alignment//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA, 2024: 25047-25057.

[96] Chen L, Chu X, Zhang X, et al. Simple baselines for image restoration// Proceedings of the European Conference on Computer Vision. Tel Aviv, Israel, 2022: 17-33.

[97] Teng M, Lou H, Yang Y, et al. Hybrid all-in-focus imaging from neuromorphic focal stack. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024, 46(12): 10124 - 10137.

[98] Duan P, Li B, Yang Y, et al. EventAid: Benchmarking event-aided image/video enhancement algorithms with real-captured hybrid dataset . IEEE Transactions on Pattern Analysis and Machine Intelligence, 2025, 47(8): 6959-6972.

[99] Li Z, Liao J, Tang C, et al. USTC-TD: A test dataset and benchmark for image and video coding in 2020s. IEEE Transactions on Multimedia, 2025, 28: 269 - 284.

[100] Xiao Z, Lu Z, Wang X. P-BiC: Ultra-high-definition image moire patterns removal via patch bilateral compensation//Proceedings of the ACM International Conference on Multimedia. Melbourne, Australia, 2024: 8365-8373.

[101] Liu R, Fan X, Zhu M, et al. Real-world underwater enhancement: Challenges, benchmarks, and solutions under natural light. IEEE Transactions on Circuits and Systems for Video Technology, 2020, 30(12): 4861-4875.

[102] Xiao Z, Li Z, Jia W. Occlusion-embedded hybrid transformer for light field super-resolution//Proceedings of the AAAI Conference on Artificial Intelligence: Vol. 39. Philadelphia, USA, 2025: 8700-8708.

[103] Xiao Z, Xiong Z. Incorporating degradation estimation in light field spatial super-resolution. Computer Vision and Image Understanding, 2025, 252: 104295.

[104] Li Z, Yuan Z, Li L, et al. Object segmentation-assisted inter prediction for versatile video coding. IEEE Transactions on Broadcasting, 2024, 70(4): 1236-1253.

[105] Zhu Y, Zhang Y, Xie X, et al. An fpga accelerator for high-speed moving objects detection and tracking with a spike camera. Neural Computation, 2022, 34(8): 1812-1839.

[106] Li Z, Li J, Li Y, et al. In-loop filtering via trained look-up tables//IEEE International Conference on Visual Communications and Image Processing. Tokyo, Japan, 2024: 1-5.

ZHAO Rui, Ph.D. His research interests include computational photography and computer vision.

XIONG Rui-Qin, Ph.D., Professor. His research interests include image and video processing, deep learning, and computational imaging.

DONG Yan-Chen, Ph.D. candidate. His research interests include computer vision, computational photography, and image restoration.

ZHAO Jing, Ph.D. Her research interests include computational photography and computer vision.

ZHENG Ya-Jing, Ph.D. Her research interests include brain-inspired computation and neuromorphic vision.

HUANG Tie-Jun, Ph.D., Professor. His main research area is visual information processing and brain-like intelligence