率失真 光学被动视觉压缩的分析

Ronald Ogden，David Fridovich-Keil

德克萨斯大学奥斯汀分校航空航天工程系，奥斯汀市威奇托街 2617 号，TX 78712

Takashi Tanaka

普渡大学航空与航天、电气与计算机工程，西体育场大道 701号，西拉法叶，印第安纳州 47907

arXiv:2602.02768v1 [cs.IT] 2 Feb 2026

摘要

利用远程视觉传感器进行自主决策，面临着如何通过资源受限信道实时传输海量视觉数据的挑战。在机器人学和控制应用中，许多系统可能迅速失稳，从而加剧这一问题，因为这需要更高的采样频率。本文提出一种新颖的感知范式：由事件相机观测视觉场景经光学生成的余弦变换，从而实现受现代视频编解码器启发的高速、无计算视频压缩。本研究对这一光学被动视觉压缩（OPVC）方案进行了仿真，并将其率失真性能与独立事件相机（SAEC）进行比较。我们发现，OPVC方案的率失真性能优于SAEC，且随着事件相机空间分辨率的提高，这一性能差距会进一步扩大。

索引词— 事件相机、傅里叶光学、率失真、压缩

1. 引言

RGB相机等视觉传感器在机器人感知与控制中发挥着核心作用。尽管视觉为自主决策提供了丰富的信息来源，但由于视觉数据具有数据量大且维度高的特点，在资源受限的通信信道上实时向远程用户传输这些传感器数据仍面临挑战。在机器人学领域，这一问题尤为突出，因为该领域通常有严格的延迟要求，例如在特技机动或高速避障场景中 [13, 14]。H.264等标准视频编解码器经过优化，能够以较高效率压缩这类密集数据，但代价是较高的计算复杂度，通常涉及复杂的运动补偿计算以及离散余弦变换（DCT）[16]。近年来，基于深度学习的自编码器也被提出作为另一种视频压缩方案 [6, 9]。然而，这类网络的一次前向传播需要经过许多卷积层，因此同样会带来不可忽视的时间复杂度。

本研究提出了一种新颖的感知范式：事件相机观测视觉场景经过光学生成的余弦变换，从而实现快速高效的编码。所提出的传感器对观测场景执行纯硬件级压缩，消除了在原始帧式视频上使用视频编解码器所涉及的计算复杂度。与许多视频编解码器一样，利用余弦变换能够根据空间频率筛选信息。事件相机编码的是场景中的变化而非完整的数据帧，因此其时间分辨率显著高于基于帧的摄像机。这使得场景能够以较低的采样率进行高效差分编码，最终或可实现对高速控制系统的闭环控制。

我们开发了一种简单、无噪声的事件相机模拟器，可处理灰度帧式视频，以分析这种新型传感器的率失真性能，并将其与独立事件相机（SAEC）的性能进行比较。我们发现，我们的光学被动视觉压缩（OPVC）方案优于 SAEC，而且随着事件相机空间分辨率的提高，性能差距也会扩大。

2. 相关工作

人们已探索并以多种形式实现光学余弦变换，通常将其用于图像压缩。Gu 等人 [8] 利用镜面系统和棱镜对输入图像进行均匀扩展，并使处理结果通过 Hughes 液晶光阀，随后使用 $4f$-系统获得相应的傅里叶变换，从而实现光学余弦变换。Wong 和 Cheng [18] 提出使用微透镜阵列和相位共轭镜，实现与 H.261 视频编解码器相匹配的光学余弦变换。Ma 等人 [11] 利用光学余弦变换辅助合成孔径雷达图像压缩。此前的这些光学余弦变换实现方式仅用于静态图像，而我们旨在用事件相机对随时间变化的场景进行编码。

关于事件相机及其应用，已有丰富的研究文献，许多内容已在全面的综述中得到总结 [2, 4]。为弥补现有事件相机数据集的不足，研究人员开发了许多事件相机模拟器。Bi 和 Andreopoulos [1] 开发了一种算法，可将基于帧的视频转换为事件流，并提供多种强度比较和时间戳确定选项。Rebecq 等人 [15]开发了一种事件相机模拟器，通过对照片级真实感渲染进行自适应采样来生成事件流；Gehrig 等人 [5] 则利用该模拟器从基于帧的视频生成事件流。Hu 等人 [10]开发了一款工具箱，可根据真实的事件相机现象（包括热像素、泄漏事件和时间噪声）从基于帧的视频生成事件流。我们实现了一种更简单但高度向量化的模拟器，可轻松集成到我们的实验框架中。

© 2026 IEEE。允许个人使用本材料。对于所有其他用途，包括在任何当前或未来的媒体中重印/再版本材料以用于广告或宣传目的、创作新的合集作品、转售或重新分发至服务器或列表，或在其他作品中重复使用本作品中任何受版权保护的部分，均须获得 IEEE 的许可。

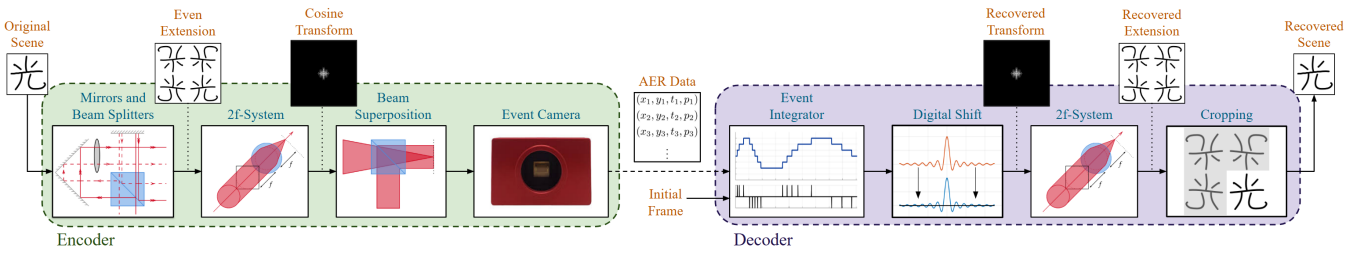

图 1。光学无源视觉压缩方案框图。

3. 压缩方法

本节介绍我们提出的压缩方法中各关键要素的背景。图 1 总结了该压缩方案。

3.1. 光学余弦变换

设函数 g 的 n维傅里叶变换 $\mathbb{R}^n \to \mathbb{C}$定义为

$$\mathscr{F}\{g\}(\boldsymbol{\omega}) := \int_{\mathbb{R}^n} g(\mathbf{x}) e^{-2\pi i \mathbf{x}\cdot\boldsymbol{\omega}}\, d\mathbf{x}. \tag{1}$$

假设在会聚透镜前方 f 处放置一块平面透明片，该透镜的焦距为 f，并用正入射的相干平面波对其进行照明。如果透明片透射光场 $U_o(\mathbf{x})$，其中$\mathbf{x} \in \mathbb{R}^2$ 是透明片处垂直于光轴的坐标平面中的一个元素，那么透镜后方 f 处的后焦平面内的光场可表示为

$$U_f(\tilde{\mathbf{x}}) = \frac{1}{i\lambda f}\mathscr{F}\{U_o\}\left(\frac{\tilde{\mathbf{x}}}{\lambda f}\right), \tag{2}$$

其中，$\tilde{\mathbf{x}} \in \mathbb{R}^2$ 表示后焦平面中的一个坐标，λ 是照明光源 [7] 的波长。虽然这使得人们能够以光学方式计算图像的傅里叶变换，但摄像机只能感知有关光场强度的信息；为了恢复原始图像，还必须获取傅里叶变换的相位。

容易证明，对于实数偶函数 g，

$$\mathscr{F}\{g\}(\boldsymbol{\omega}) = \int_{\mathbb{R}^n} g(\mathbf{x}) \prod_{k=1}^{n} \cos(2\pi x_k \omega_k)\, d\mathbf{x} := \mathscr{F}_c\{g\}(\boldsymbol{\omega}), \tag{3}$$

其中，粗体变量表示 $\mathbb{R}^n$ 中的坐标，x_k 和 ω_k 分别表示$\mathbf{x}$ 和 $\boldsymbol{\omega}$ 的 k^{th} 个元素。我们将 $\mathscr{F}_c\{g\}$ 称为 g的余弦变换。因此，实数偶函数的傅里叶变换是实值的，且等价于该函数的余弦变换。任何具有紧支撑的实函数都可以进行偶延拓，从而能够根据其偶延拓唯一恢复该函数。在光学上，可以使用镜面系统和分束器 [8]实现这一点。

所得偶延拓的傅里叶变换为实数，因此只需进行二元相位恢复。在光学上，可通过将平面波 $U_r(\tilde{\mathbf{x}}) = A_r \in \mathbb{R}_{>0}$ 叠加到傅里叶平面上来实现，使得 $A_r > \max_{\tilde{\mathbf{x}}} |U_f(\tilde{\mathbf{x}})|^2$，如 [11, 18] $I_m \tilde{} = |U_f \tilde{} + $ 中所示。给定强度 $A_r|^2$ A_r 测量 (x) (x)，假设已知，便可使用 $U_f \tilde{} = \sqrt{I_m \tilde{} - A_r}$ A_r $U_f \tilde{} + A_r >$ (x) (x) 恢复变换，因 0 为我们选择了使得 (x) 的方式。

3.2. 事件相机

事件相机由异步工作的像素阵列组成。只有当像素所感知光线的对数强度相较于上一次事件发生时的偏差超过某个阈值，该像素才会触发事件。也就是说，假设某个像素上一次事件发生在时间 $t_{k-1} \in \mathbb{R}$，且该像素处光线的对数强度随时间变化为$L(t) := \log(I(t))$，其中 $I = |U|^2$ 使用我们的记法，见第 3.1 节。那么，该像素下一次事件发生的时间为

$$t_k := \inf\{t : t > t_{k-1}, |L(t) - L(t_{k-1})| \geq \delta\}, \tag{4}$$

其中，δ 表示为传感器设定的对称事件阈值。在时间 t_k，该像素发送以下地址事件表示（AER）数据包：$(\mathbf{u}_k,\ p_k,\ t_k)$，其中$\mathbf{u}_k$ 是像素位置，$p_k \in \{1, -1\}$ 是变化的极性，其定义为

$$p_k := \operatorname{sgn}\left(L(t_k) - L(t_{k-1})\right). \tag{5}$$

在所提出的压缩方案中，编码器将这段 AER数据流发送给解码器，由事件积分器对事件相机记录的场景进行有损重建。给定事件时间序列 $\{t_k\}_{k=1}^K$ 及对应的事件极性序列 $\{p_k\}_{k=1}^K$，对于初始感知对数强度已知为 $L(0)$的像素，可以按如下方式对事件流进行积分，从而生成该像素对数强度历史的有损重建：

$$\tilde{L}(t) = L(0) + \sum_{k:t_k<t} \delta p_k. \tag{6}$$

在实际应用中，如果未能以足够高的频率提供完整的参考帧，这种简单的重建方法对传感器噪声缺乏鲁棒性。不过，我们不考虑传感器噪声，因此这种简单的重建方法可以避免去噪滤波器可能给研究引入的混杂影响。请注意，随着事件阈值 δ 增大，事件发生的频率会降低，但 $L(t)$ 与 $\tilde{L}(t)$ 之间的差异会增大。

在本研究中，我们假设解码器已知每个像素的初始强度。实际操作中，可以在光学装置中加入低帧率帧式相机，为解码器提供周期性的参考帧，从而也能增强其对噪声的鲁棒性。

4. 计算分析

我们新颖的 OPVC方案与传统视频编解码器不可相提并论；前者旨在对连续场景进行无需计算的压缩，而后者旨在优化基于帧的压缩。作为 OPVC的替代方案，可以使用独立事件相机直接观测场景。为了使用 SAEC对视频进行编码/解码，计算资源仅

用于事件积分。对于 OPVC，光学元件以被动方式计算余弦变换，因此除事件积分外，在编码/解码期间仅需通过计算处理少量简单的逐元素运算。因此，就每个事件而言，两种压缩方案具有相同数量级的计算复杂度，并且都对连续场景执行硬件级压缩，使其具有可比性。

4.1. 率失真框架

事件相机将连续的视觉场景转换为离散的 AER 元组列表，从而产生失真的重建结果。这类有损压缩方法适合在率失真框架中进行比较，该框架量化了重建质量（失真）与随时间用于编码视频的比特数（码率）之间的权衡 [3]。对于给定的事件相机和时间分辨率，每个 AER 元组占用相同数量的比特进行存储，因此在此分析中，我们使用采样率而非比特率。尽管有多种失真度量适用于此应用，我们采用多尺度结构相似性（MS-SSIM） [17]。

将这一思路精确表述如下：令 $I(t) \in \mathbb{R}_{\geq 0}^{m\times n}$ 表示视觉场景在时间 $t \in \mathbb{R}$ 的、大小为 $m \times n$ 的像素强度图像。对于此次比较，给定一个场景分布 $\mathbb{P}$ 及时间范围T，我们希望计算如下定义的率失真曲线：

$$\begin{aligned} R(D;\pi) &:= \inf_{\delta} \mathbb{E}_{I\sim\mathbb{P}}\left[\frac{K}{mnT}\right], \\ \text{s.t.}\quad & \mathbb{E}_{I\sim\mathbb{P}}\left[\frac{1}{T}\int_0^T \operatorname{dist}\left(I(t),\tilde{I}(t)\right)\,dt\right] \leq D, \\ & \tilde{I} = \operatorname{dec}_{\pi(\delta)}(\operatorname{enc}_{\pi(\delta)}(I)), \end{aligned} \tag{7}$$

其中， R 是采样率， D 是失真， π 表示压缩方法（例如 SAEC 或 OPVC），该方法由事件阈值 δ参数化；$\operatorname{enc}_{\pi(\delta)}(\cdot)$ 和 $\operatorname{dec}_{\pi(\delta)}(\cdot)$ 分别是相应的编码函数和解码函数； K 是截至时间 T发生的事件数量；$\operatorname{dist}(\cdot)$ 是我们选用的失真度量（ MS-SSIM）； $\tilde{I}$ 表示重建场景。

4.2. 仿真方法

为了分析 SAEC 和 OPVC 的率失真性能，我们实现了一个简单的事件相机模拟器。模拟器的输入为未经压缩的灰度基于帧的视频，由帧 $I(s) \in \mathbb{R}_{\geq 0}^{m\times n}$ 组成，其中 $s \in [S]$表示帧索引， $\delta \in \mathbb{R}_{>0}$表示事件阈值。模拟器输出事件计数 K，对应于编码器发送给解码器的 AER数据包总数，以及由解码器生成的重建视频 $\tilde{I}(s)$。虽然式（7）将率失真问题表述为连续时间问题，但我们将调整处理方式，以适应仿真的离散特性。

模拟器对事件的计数方式，假设原始视频帧所表示的连续时间场景的对数强度在帧间单调且连续地变化。我们假设每个像素的时间分辨率和不应期都足够小，因此像素不会“漏掉”任何事件。这些假设得出的事件计数方式与 [10] 中的事件生成模型相似。当原始基于帧的视频中的物体移动缓慢时，此方法的准确性会提高。

该模拟假设解码器可以访问初始帧 $I(0)$，因此设置 $\tilde{I}(0) = I(0)$。模拟器通过计算 $L(s)$（即对 $I(s)$ 逐元素取对数），将帧转换到对数空间。解码器还会跟踪对数空间重建 $\tilde{L}(s)$，其定义如下假设解码器已重建帧 $\tilde{L}(s-1)$。根据像素的对数强度在帧间单调变化这一假设，特定像素在帧间感知到的所有事件都具有相同的极性。令 $E\colon [S]\setminus\{0\} \to \mathbb{Z}^{m\times n}$表示一个事件函数，用于表示帧 $s-1$ 之后、帧 s 之前每个像素发生的事件数量乘以各自极性。该函数可按如下方式计算：

$$E(s) = \left\lceil \frac{L(s) - \tilde{L}(s-1)}{\delta} \right\rfloor, \tag{8}$$

其中， $\lceil\cdot\rfloor$ 表示逐元素向零舍入。该矩阵编码了帧 $s-1$ 和 s 之间发生的所有事件的极性和像素位置信息。随后，解码器将 $\tilde{L}(s)$ 重建为 $\tilde{L}(s) = \tilde{L}(s-1) + \delta E(s)$。

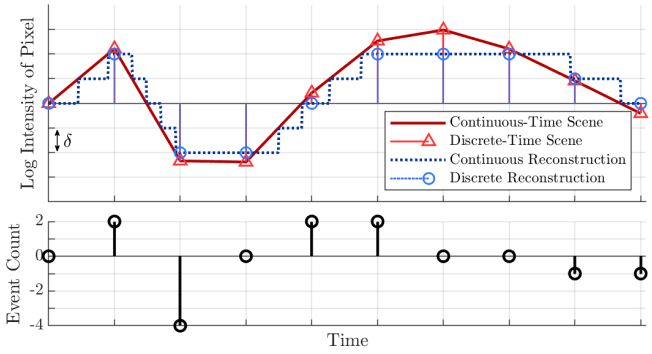

图 2。仿真方法示意图。红色三角形表示原始视频数据的对数强度， $L(s)$。红色实线表示假设的帧间连续像素变化，蓝色虚线表示相应的连续重建。蓝色圆圈表示基于帧的重建， $\tilde{L}(s)$。下方图示表示式（8）中定义的事件函数 $E(s)$。

该仿真方法如图 2 所示。令$E_{ij}(s)$表示 $E(s)$中第 i行、第 j 列的元素，则仿真期间的事件总数为

$$K = \sum_{i=0}^{m-1}\sum_{j=0}^{n-1}\sum_{s=0}^{S-1} |E_{ij}(s)|\,. \tag{9}$$

最后一步，模拟器将重建的对数强度帧转换回强度帧。该模拟器可以分析 SAEC 的性能，但 OPVC 还需要额外的组件。

对于 OPVC，编码器使用二维离散余弦变换来模拟光学余弦变换。若 $\mathcal{I}(s) \in \mathbb{R}^{m\times n}$ 表示原始视频中 $I(s)$的余弦变换，则

$$\mathcal{I}_{qr}(s) = \alpha_q\alpha_r \sum_{i=0}^{m-1}\sum_{j=0}^{n-1} I_{ij}(s)\cos\frac{\pi(2i+1)q}{2m}\cos\frac{\pi(2j+1)r}{2n}, \tag{10}$$

其中， $\alpha_q = \sqrt{1/m}$ 对于 $q = 0$， $\alpha_q = \sqrt{2/m}$ 对于 $q > 0$，而 α_r 的定义与此类似。OPVC 模拟器会将变换后的视频整体加上一个足够大的正常数偏移，使其在传递给模拟事件相机之前，所有帧中的每个像素值都为正。解码器会从事件相机的重建输出中减去相同的常数。这些常数偏移

模拟了第 3.1 节中讨论的用于相位恢复的平面波叠加。最后，我们对结果进行逆离散余弦变换，得到最终重建结果 $\tilde{I}(s)$。

给定单个视频，该模拟会计算以下采样率和失真：

$$R_{\text{sim}}(I;\pi(\delta)) = K/(mnS), \tag{11}$$

$$D_{\text{sim}}(I;\pi(\delta)) = \frac{1}{S}\sum_{s=0}^{S-1} \text{MS-SSIM}\left(I(s), \tilde{I}(s)\right). \tag{12}$$

对于每种压缩方法，我们计算了多个视频中 R_{sim} 和 D_{sim} 的平均值。我们针对一系列事件阈值重复这些计算，以构建率失真曲线。我们通过逐块像素平均对数据集进行空间下采样，使其达到所需分辨率，从而在不同视频分辨率下开展了这项分析。我们使用 Ultra Video Group (UVG) 数据集进行分析 [12]。该数据集旨在用于视频编解码器分析，包含 16 个分辨率为 3840×2160、最长达 600 帧的未压缩视频。

4.3. 结果

图 3 展示了第 4.2 节所述模拟的结果。在测试的每种分辨率下，对于任意给定的失真约束，OPVC 的每像素采样率都低于 SAEC。此外，随着事件相机分辨率的提高，OPVC 与 SAEC 之间的每像素性能差距也会扩大。空间下采样会固有地滤除观测场景中的高空间频率信息，因此可能会削弱 OPVC 相较于 SAEC 在较低分辨率下的优势。我们还使用均方误差、结构相似性指数和峰值信噪比等不同的失真度量进行了类似实验；对于该数据集，所有度量都呈现出相同的定性结果。

图 4 展示了 SAEC 和 OPVC 仿真的定性比较。SAEC 似乎保留了原始场景中的许多清晰边缘，但由于高对比度物体之间的边缘运动，产生了失真伪影。OPVC 重建结果似乎没有这些运动伪影，但存在高频噪声，使原始场景中清晰的边缘变得模糊。这是因为光学余弦变换将图像的大部分能量集中到低频域，导致事件相机滤除了高频信息。

5. 结论

在本研究中，我们提出了一种高速、计算量极低的压缩方案，将傅里叶光学与事件相机相结合，以被动滤除视觉场景中的高空间频率信息。对所提出压缩方法的仿真表明，从率失真角度来看，该方法优于独立事件相机。尽管这些初步结果展现出良好前景，但它们代表的是一种理想情形。未来的研究应纳入事件相机传感器噪声、更高保真度的基于帧的视频到事件数据的转换（见 [1], [10]和 [19]），并考虑加入周期性初始化帧后产生的率失真权衡。

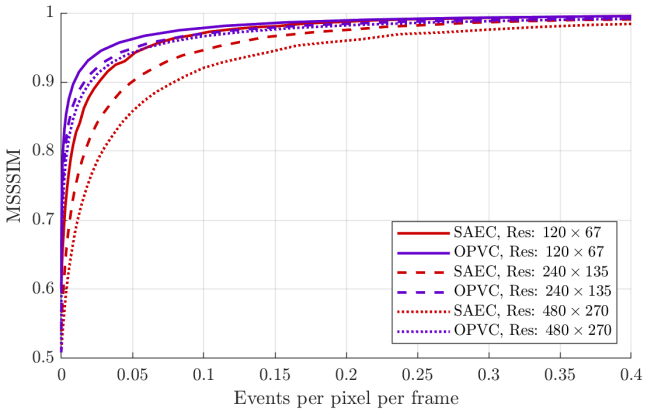

图 3。SAEC 和 OPVC 在 UVG 数据集不同空间分辨率下的率失真性能。



图 4。压缩结果的定性比较。顶部图像是原始视频中的一帧，画面中一名骑师骑着马向左行进，摄像机随之摇摄。中间图像是使用 SAEC 以 0.9 MS-SSIM 重建的一帧。底部图像是使用 OPVC 以 0.9 MS-SSIM 重建的一帧。

参考文献

[1] Yin Bi和 Yiannis Aloimonos。将像素域视频帧参数化转换为神经形态视觉流。载于*2017*年*IEEE*国际图像处理会议 *(ICIP)*，第1990–1994 页。IEEE，2017年。

[2] Bharatesh Chakravarthi、Aayush Atul Verma、Kostas Dani-ilidis、Cornelia Fermuller 和 Yezhou Yang。近期事件相机创新：综述。载于欧洲计算机视觉会议，第 342–376 页。施普林格，2024年。

[3] Thomas M Cover 约翰·威立父子出版公司，1999年。

[4] Guillermo Gallego, Tobi Delbr¨uck, Garrick Orchard, Chiara Bartolozzi, Brian Taba, Andrea Censi, Stefan Leutenegger, An- drew J Davison, J¨org Conradt 等。事件视觉：综述。*IEEE*模式分析与机器智能汇刊，44(1):154–180，2020。

[5] Daniel Gehrig、Mathias Gehrig、Javier Hidalgo-Carri´o 和 Davide Scaramuzza。视频转事件：将视频数据集用于事件相机。在*IEEE/CVF*计算机视觉与模式识别会议论文集，第3586–3595页，2020年。[6] Adam Golinski、Reza Pourreza、Yang Yang、Guillaume Sautiere 和 Taco S Cohen。用于视频压缩的反馈循环自编码器。在亚洲计算机视觉会议论文集，2020年。[7] Joseph W Goodman。傅里叶光学导论。罗伯茨公司出版社，2005年。[8] Zu-Han Gu、James R Leger 和 Sing H Lee。余弦变换的光学计算。《光学通讯》，39(3)：137–142，1981年。[9] Amirhossein Habibian、Ties van Rozendaal、Jakub M Tom- czak 和 Taco S Cohen。基于率失真自编码器的视频压缩。在*IEEE/CVF*国际计算机视觉会议论文集，第7033–7042页，2019年。[10] Yuhuang Hu、Shih-Chii Liu 和 Tobi Delbruck。v2e：从视频帧生成逼真的DVS事件。在*IEEE/CVF*计算机视觉与模式识别会议论文集，第1312–1321页，2021年。[11] Jiacheng Ma、Baozhang Yang、Yesheng Gao、Lei Tao 和 Xingzhao Liu。利用光电子处理进行合成孔径雷达图像压缩。《工程学报》，2019(19)：5982–5985，2019年。[12] Alexandre Mercat、Marko Viitanen 和 Jarno Vanne。UVG数据集：用于视频编解码器分析与开发的50/120 fps 4K序列。在第*11*届*ACM*多媒体系统会议论文集，第297–302页，2020年。[13] Elias Mueggler、Basil Huber 和 Davide Scaramuzza。面向高速机动的基于事件的六自由度位姿跟踪。在*2014*年*IEEE/RSJ*智能机器人与系统国际会议，第2761–2768页。IEEE，2014年。

[14] Elias Mueggler、Nathan Baumli、Flavio Fontana 和 Davide Scaramuzza。利用动态视觉传感器实现四旋翼飞行器的规避机动。在*2015*年欧洲移动机器人会议（*ECMR*），第1–8页。IEEE，2015年。[15] Henri Rebecq、Daniel Gehrig 和 Davide Scaramuzza。Esim：一款开放式事件相机模拟器。在机器人学习会议，第969–982页。PMLR，2018年。[16] Khalid Sayood。数据压缩导论。Morgan Kaufmann，2017年。[17] Zhou Wang、Eero P Simoncelli 和 Alan C Bovik。用于图像质量评估的多尺度结构相似性。在*2003*年第三十七届阿西洛马会议：信号、系统与计算机，第2卷，第1398–1402页。IEEE，2003年。[18] KW Wong KW Wong 和 LM Cheng LM Cheng。利用微透镜阵列和相位共轭镜实现光学余弦变换。《日本应用物理学杂志》，31(5S)：1672，1992年。[19] Zhongyang Zhang、Shuyang Cui、Kaidong Chai、Haowen Yu、Subhasis Dasgupta、Upal Mahbub 和 Tauhidur Rahman。V2ce：视频到连续事件模拟器。在*2024*年*IEEE*国际机器人与自动化会议（*ICRA*），第12455–12461页。IEEE，2024年。