2026 IEEE International Conference on Robotics and Automation (ICRA 2026)
June 1-5, 2026. Vienna, Austria

Efficient Event Camera Volume System

Juan Soto[1*], Ian Noronha[1*], Saru Bharti[1], Upinder Kaur[1†]

***Abstract*— Event cameras promise low latency and high dynamic range, yet their sparse output challenges integration into standard robotic pipelines. We introduce EECVS (Efficient Event Camera Volume System), a novel framework that models event streams as continuous-time Dirac impulse trains, enabling artifact-free compression through direct transform evaluation at event timestamps. Our key innovation combines density-driven adaptive selection among DCT, DTFT, and DWT transforms with transform-specific coefficient pruning strategies tailored to each domain's sparsity characteristics. The framework eliminates temporal binning artifacts while automatically adapting compression strategies based on real-time event density analysis. On EHPT-XC and MVSEC datasets, our framework achieves superior reconstruction fidelity with DTFT delivering the lowest earth mover distance. In downstream segmentation tasks, EECVS demonstrates robust generalization. Notably, our approach demonstrates exceptional cross-dataset generalization: when evaluated with EventSAM segmentation, EECVS achieves mean IoU 0.87 on MVSEC versus 0.44 for voxel grids at 24 channels, while remaining competitive on EHPT-XC. Our ROS2 implementation provides real-time deployment with DCT processing achieving 1.5 ms latency and 2.7× higher throughput than alternative transforms, establishing the first adaptive event compression framework that maintains both computational efficiency and superior generalization across diverse robotic scenarios.**

I. Introduction

Robotic systems are increasingly required to operate in environments that place strong demands on perception, such as high-speed navigation, low-light conditions, or scenes with extreme contrast. Standard RGB cameras suffer fundamental limitations in these scenarios: motion blur corrupts fast movements, limited dynamic range loses critical information, and high processing overheads constrain real-time performance [1] [2]. These constraints fundamentally limit robotic perception reliability and restrict autonomous operation in dynamic environments.

Event cameras address these limitations through asynchronous, pixel-independent brightness change detection. Unlike traditional cameras capturing full frames at fixed intervals, event cameras generate continuous streams of spatiotemporal events containing coordinates, timestamps, and polarity information. This sensing paradigm enables key properties, such as microsecond temporal resolution, high dynamic range of up to 140 dB, and low power consumption [3]. These characteristics make event cameras particularly suitable for robotic applications that require fast and reliable perception in visually demanding scenarios with limited computational resources [3].

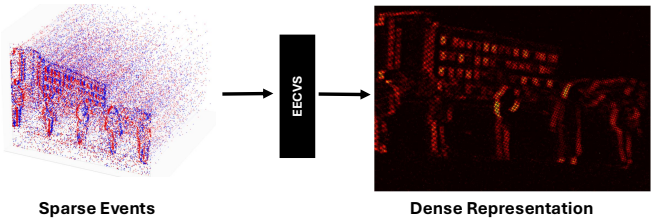

Fig. 1. **Event-to-dense representation in EECVS.** Incoming event streams are processed within the framework and converted into compact dense representations through the application of DCT, DTFT, or DWT.

Despite these advantages, the sparse and asynchronous nature of event data creates a fundamental integration challenge: standard vision pipelines expect dense inputs, while events arrive as irregular spatiotemporal streams. Current approaches employ fixed representations that ignore stream variability. Others convert sparse events into dense images for efficient processing, but sacrifice fine temporal resolution [4]. Voxelized representations preserve spatiotemporal structure yet blur information through temporal binning. TORE volumes encode only recent events, discarding longer temporal dependencies [5]. CES volumes apply single transforms, such as DFT (Discrete Fourier Transform), for compact encodings, but lack the flexibility to adapt across diverse scenarios [6]. These highlight the critical need for adaptive compression frameworks that can intelligently respond to varying event stream characteristics.

We address this challenge through EECVS , a unified framework that adaptively compresses event streams using autonomous density-driven transform selection. Our framework leverages compressed sensing theory, which proves that sparse signals in appropriate domains require few coefficients for accurate reconstruction. EECVS exploits event stream sparsity through intelligent, density-aware transform selection for real-time adaptation to changing scene dynamics. As illustrated in Fig. 1, the framework ingests event streams and outputs dense representations that serve as the basis for robotic perception. The framework adaptively chooses among three complementary transforms: the **Discrete Cosine Transform (DCT)** excels at energy concentration in high-density regions, the **Discrete Time Fourier Transform (DTFT)** preserves temporal fidelity for moderate event rates, and the **Discrete Wavelet Transform (DWT)** maintains localized structure for sparse activity patterns. Our approach models event streams as continuous-time Dirac impulse trains, enabling direct transform evaluation at event

[1]Purdue University, USA. {inoronha, soto97, bharti3, kauru}@purdue.edu
*These authors contributed equally to this work. † Corresponding author.
Project page and code: https://github.com/Dookiep/EECVS.git.

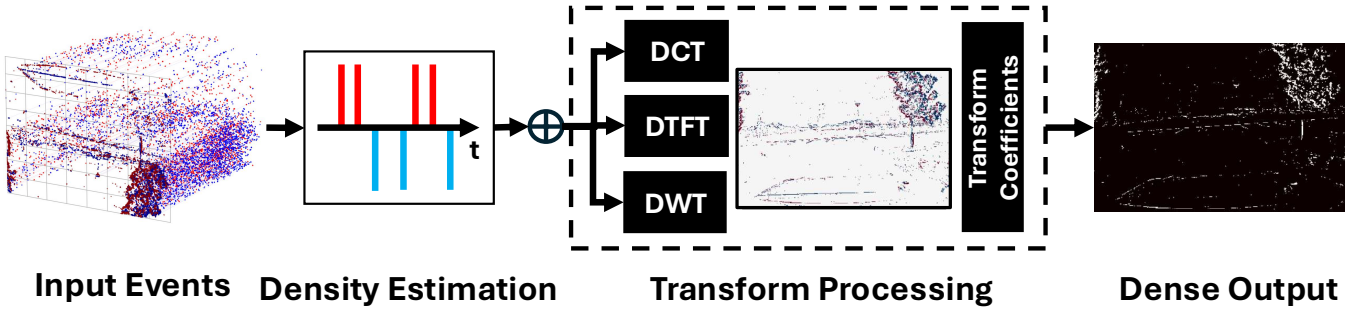

Fig. 2. Compression process for a single event window. Events are aggregated, transformed with a basis selected according to activity density, pruned by either low-frequency retention (DCT) or magnitude selection (DTFT/DWT), and packed into dense representations.

timestamps without temporal binning artifacts that compromise reconstruction quality. This mathematical foundation, combined with density-driven selection criteria grounded in information-theoretic principles, produces dense representations that minimize information loss while ensuring compatibility with robotic perception pipelines.

In this paper, we address the need for efficient and adaptable event representations in robotics and present a unified compression framework. Our contributions are:

- **EECVS autonomous compression framework:** An intelligent framework that combines continuous-time Dirac impulse modeling with density-driven transform selection, eliminating temporal binning artifacts while enabling task-adaptive performance across diverse robotic scenarios with computational efficiency.
- **Systematic Reconstruction Analysis:** Comprehensive evaluation demonstrating how each transform preserves temporal fidelity, spatial detail, and information content under varying event densities, providing insights for optimal compression strategy selection.
- **Downstream Validation and Real-world Deployment:** Extensive downstream task evaluation revealing superior generalization capabilities, coupled with a ROS2 implementation supporting real-world robotic deployment.

II. Related Work

Early event processing methods faced fundamental trade-offs between temporal precision and computational tractability. Sequential processing approaches achieve microsecond latency through probabilistic filters and spiking neural models [7], [8]. These techniques update state sequentially using continuous time dynamics, but their reliance on handcrafted rules and parameter tuning limits their scalability to higher-level perception tasks. Spiking Neural Networks (SNN) extend this paradigm with data-driven approaches [9], [10], but practical use remains difficult due to training instabilities.

Aggregation-based methods trade temporal precision for computational efficiency and compatibility with standard vision pipelines. Time Ordered Recent Event (TORE) [5] representations store short temporal histories in timestamp grids to reduce update costs and capture local temporal sequences. However, it still quantizes past information and loses fine detail, thereby eroding the benefits of using event cameras. Other approaches aggregate events into grid or voxel-based structures that accumulate event counts or polarities over fixed intervals, producing dense representations that are compatible with convolutional networks [11]–[13]. These methods intuitively highlight brightness changes and edges but inevitably discard fine timing by merging events into coarse temporal bins.

Recent research explores improved event representations and learning-based processing strategies. ERGO-12 [14] studies how the choice of dense event encodings affects downstream learning performance and proposes selecting representations using the Gromov–Wasserstein discrepancy. Other approaches address the variability of event density through adaptive stacking strategies, such as multi-density event stacks (MDES) [15]. Event focal stack representations [16] further exploit the temporal continuity of events to encode scene structure across time. In parallel, other works represent events as sets of spatiotemporal points processed by graph networks or transformers [17]–[19], preserving sparsity and temporal resolution. However, these approaches increase computational overheads and are often tailored toward specific perception tasks.

Compressed sensing provides the theoretical foundation for our framework, adapting compression strategies across multiple domains. Medical imaging pioneered adaptive compression through Sparse MRI, which dynamically adjusts acquisition strategies [20]. Computed tomography employs compressed sensing for adaptive low-dose reconstruction [21]. Seismic imaging in geophysics applies scene-dependent processing for subsurface recovery [22]. These successes demonstrate that adaptive selection among transforms enhances compression effectiveness when matched to signal characteristics.

Event data exhibits natural spatiotemporal sparsity, making it ideal for adaptive compression frameworks. However, existing approaches lack principled selection criteria for choosing among transform bases. Our framework addresses this gap by establishing density-driven selection criteria grounded in compressed sensing theory, enabling adaptive compression that responds to varying event stream characteristics.

III. Methods

EECVS adaptively compresses event streams through density-driven transform selection and coefficient pruning strategies tailored to each transform's characteristics. The framework continuously monitors activity levels within temporal windows. This real-time analysis drives autonomous selection among three specialized transforms. We derive an event density measure that helps classify the perceived event streams. For sparse windows, the framework deploys wavelets to preserve isolated features. When detecting moderate activity, EECVS switches to Fourier analysis for temporal fidelity. Dense activity triggers the framework's cosine transform for maximum energy compaction. Following transform selection, EECVS autonomously reduces coefficients to a proportion $r = \min|M, \mathcal{K}_w|$, using transform-specific pruning strategies. For DCT, EECVS retains the first r low-frequency indices. For DTFT and DWT, the system selects the r largest-magnitude coefficients. The framework then packages these compact descriptors into dense representations compatible with standard perception pipelines (Fig. 2).

A. Event Density Estimation and Transform Selection

To adaptively compress event streams, we introduce an event density measure that quantifies the activity levels within temporal windows and provides the basis for principled transform selection. Let the input window be

$$E_w = \{(t_i, x_i, y_i, p_i)\}_{i=1}^{N_w},$$

where $N_w = |E_w|$ is the number of events observed in window w, t_i is the timestamp, (x_i, y_i) are pixel coordinates and $p_i \in \{-1, +1\}$ is the polarity. For a sensor of height H and width W, and a window duration T, the normalized event density is defined as

$$\rho_w = \frac{|E_w|}{H \cdot W \cdot T}, \tag{1}$$

expressed events per pixel per unit time (events/s). This normalization allows consistent comparisons across different sensor resolutions and temporal scales.

We establish three density regimes based on empirical analysis of event distribution characteristics. Our threshold selection uses percentile-based partitioning, which adapts naturally to sensor and scene characteristics while providing stable selection criteria. We partition event streams according to thresholds τ_{low} and τ_{high}:

$$\text{Regime}(w) = \begin{cases} \text{Sparse}, & \rho_w < \tau_{\text{low}}, \\ \text{Moderate}, & \tau_{\text{low}} \le \rho_w < \tau_{\text{high}}, \\ \text{Dense}, & \rho_w \ge \tau_{\text{high}}. \end{cases} \tag{2}$$

Thresholds τ_{low} and τ_{high} are set on a calibration split using percentiles of the window density ρ_w in (1). We compute the empirical distribution of ρ_w over all windows in the split and define τ_{low} as the 25th percentile and τ_{high} as the 75th percentile. This choice ensures balanced workload distribution while maintaining theoretical grounding: the lowest quartile typically represents sparse, isolated events best suited for localized wavelet representation; the highest quartile contains dense activity requiring energy concentration through cosine transforms; the middle half exhibits moderate structure optimally preserved by Fourier analysis. With this choice, windows in the lowest quartile are assigned to DWT, windows in the highest quartile are assigned to DCT, and the middle half is assigned to DTFT, consistent with (2) and (3). Percentile thresholds adapt to the sensor and scene without manual tuning, since ρ_w is normalized by H, W, and T. After calibration, thresholds remain fixed during deployment to ensure consistent selection behavior.

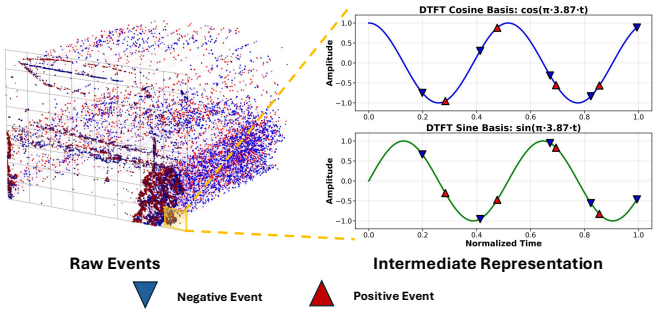

Fig. 3. Per pixel DTFT within a window. Events sample transform bases directly, and coefficients are pruned according to the transform-specific retention strategy.

For each density regime, the framework selects the transform domain that yields the sparsest and most informative representation:

$$\text{Transform}(w) = \begin{cases} \text{DWT}, & \rho_w < \tau_{\text{low}}, \\ \text{DTFT}, & \tau_{\text{low}} \le \rho_w < \tau_{\text{high}}, \\ \text{DCT}, & \rho_w \ge \tau_{\text{high}}. \end{cases} \tag{3}$$

This selection strategy exploits the complementary sparsity of DWT, DTFT, and DCT across density regimes, forming the basis for event-driven transform encoding.

B. Event-Driven Signal Model and Transform Encoding

a) Dirac impulse model.: Within a window w, we model the event stream as a sum of Dirac impulses in continuous time, avoiding temporal binning artifacts:

$$s_w(t) \;=\; \sum_{i=1}^{N_w} p_i \, \delta(t - t_i)\,, \tag{4}$$

where each event (t_i, x_i, y_i, p_i) contributes polarity $p_i \in \{-1, +1\}$ at timestamp t_i. This formulation enables direct evaluation of transform atoms at event timestamps, preserving temporal precision while maintaining computational efficiency (see Fig. 3). EECVS processes each pixel independently: for every spatial location (x, y) we accumulate basis samples from that pixel's events within the current window. The transform choice derives from window-level density analysis, while coefficient computation occurs at the pixel level.

b) Coefficient accumulation by basis sampling.: Let $\{\phi_k(t)\}_{k \in \mathcal{K}_w}$ denote the transform atoms selected for window w by the density-driven selection rule (Section above,

cf. Eq. (1)). Using the distributional property of the Dirac delta function δ, we compute coefficients as:

$$\begin{aligned} c_{w,k} &= \langle s_w, \phi_k \rangle \\ &= \int s_w(t)\, \phi_k(t)\, dt \\ &= \sum_{i=1}^{N_w} p_i\, \phi_k(t_i), \quad k \in \mathcal{K}_w. \end{aligned} \tag{5}$$

We order $\mathcal{K}_w$ to prioritize low frequencies or coarse scales, which concentrate energy and promote sparsity in the compressed representation.

c) Transforms by regime.: Given $\mathrm{Transform}(w) \in \{\mathrm{DWT}, \mathrm{DTFT}, \mathrm{DCT}\}$, each transform provides distinct advantages:

- **Sparse regime (DWT).** Wavelet atoms $\psi_{j,m}(t)$ at coarse scales j and shifts m provide localized representation ideal for isolated events: $c_{w,(j,m)} = \sum_i p_i\, \psi_{j,m}(t_i)$, with $(j,m) \in \mathcal{K}_w$ biased toward coarser scales to preserve isolated activity structure.
- **Moderate regime (DTFT).** Complex sinusoids $\phi_\omega(t) = e^{-\,\mathrm{i}\omega t}$ over a finite set of low angular frequencies $\omega \in \Omega_w$ maintain temporal fidelity: $c_{w,\omega} = \sum_i p_i\, e^{-\,\mathrm{i}\omega t_i}$. We retain the lowest $|\omega|$ values to preserve temporal structure with compact spectral representation.
- **Dense regime (DCT).** Real cosine atoms $\phi_k(t) = \cos(\omega_k t)$ with ordered frequencies ω_k provide efficient energy compaction: $c_{w,k} = \sum_i p_i \cos(\omega_k t_i)$, keeping the lowest indices $k \in \mathcal{K}_w$ for strong energy concentration in dense activity patterns.

C. Coefficient Retention and Representation Packing

We use a fixed coefficient budget of M per window. Unless stated otherwise, we use $M = 16$, as it provides a balance between compression efficiency and reconstruction quality. For each window, we retain

$$r = \min\{M,\, |\mathcal{K}_w|\},$$

ensuring that the number of selected coefficients does not exceed the available basis size. With this budget established, coefficient pruning strategies match transform characteristics:

For **DCT**, we use frequency retention. We order cosine indices from low to high and keep the first r. This promotes energy compaction in dense activity. For **DTFT** and **DWT**, we conduct magnitude selection. We rank all coefficients by $|c_{w,k}|$ over the chosen frequency grid (DTFT) or over the scale and shift atoms (DWT) and keep the top r. This is robust in sparse or moderately populated windows where energy is not localized at low frequency. Ties are resolved in favor of the lower frequency or coarser scale.

The packed descriptor is $\hat{\mathbf{c}}_w = (c_{w,k})_{k \in \mathcal{K}_w^\alpha}$, with $\mathcal{K}_w^\alpha$ equal to the low frequency set for DCT and the top magnitude set for DTFT or DWT. We also report ablations with $M = 8$ and $M = 24$ to study the trade off between bitrate and fidelity. Windows are processed independently and then concatenated for the downstream perception stack. The effect of M on reconstruction is summarized in Table I.

TABLE I

OVERALL MEANS FOR TWO COEFFICIENT SIZES USING SPATIOTEMPORAL METRICS (MSE↓, SSIM↑, EMD↓).

Method	M	MSE ($\times 10^{-3}$)	SSIM	EMD
DCT	8	**1.731**	**0.850**	0.964
	24	1.740	0.849	0.964
DTFT	8	1.389	0.882	0.964
	24	**1.350**	**0.887**	0.964
DWT	8	**2.010**	0.819	0.964
	24	2.011	0.819	0.964

D. ROS2 implementation

We implemented EECVS as a modular ROS2 framework for seamless real-time robotic deployment. The system comprises three autonomous components operating together to provide comprehensive event stream processing. First, an event emulator generates configurable synthetic streams, enabling reproducible testing without specialized hardware. Second, a density estimation module continuously monitors event distributions and publishes real-time metrics that drive transform selection. Third, a logging module records compression features and coefficients, enabling offline analysis of the framework's autonomous decisions.

The implementation follows ROS2 conventions to ensure integration compatibility with existing robotic perception stacks. EECVS's modular architecture supports substitution of compression operators. Developers can extend the framework with additional transforms while preserving autonomous selection. This design enables flexible deployment across simulation and physical robots.

IV. EXPERIMENTS AND RESULTS

A. Experimental Setup

We evaluate EECVS against established baselines to demonstrate autonomous compression effectiveness across diverse scenarios. Our framework competes with CES volumes and TORE volumes, representing current state-of-the-art approaches to event compression and representation. For EECVS evaluation, we configure $M = 16$ retained coefficients.

Experiments span EHPT-XC [23] and MVSEC [24] datasets. These datasets challenge the framework with significantly different density characteristics (Fig. 4), with EHPT-XC containing high-density real-world sequences and MVSEC providing moderate motion-driven densities. Nine representative scenarios test the framework's autonomous adaptation across indoor/outdoor environments, varying lighting conditions, and different motion patterns. This evaluation design reveals how EECVS balances compression efficiency, reconstruction quality, and computational robustness.

B. Reconstruction Performance Analysis

Reconstruction metrics quantify EECVS's information preservation capabilities through adaptive selection of compression schemes. The framework reconstructs signals by

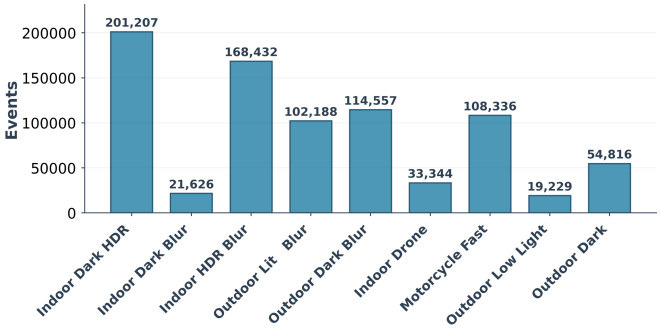

Fig. 4. Histogram of event densities across the EHPT-XC and MVSEC datasets. EHPT-XC exhibits high-density real-world sequences, while MVSEC contains more moderate motion-driven densities.

inverse-transforming retained coefficients, enabling direct comparison with original event streams. We measure Mean Squared Error (MSE) and Structural Similarity Index (SSIM) on rendered frames, while Earth Mover Distance (EMD) evaluates temporal alignment preservation.

Table II demonstrates that EECVS's DTFT selection delivers superior performance across diverse scenarios. The framework achieves the smallest EMD in eight of nine experiments while maintaining the lowest MSE and highest SSIM values. In Indoor (Dark+HDR), DTFT achieves remarkable reconstruction fidelity with MSE 0.0001, SSIM 0.987, and EMD 1.102. For Indoor (HDR+Blur), EECVS again delivers exceptional performance with MSE 0.0001 and EMD 0.642, significantly outperforming alternative approaches.

The framework demonstrates intelligent adaptation to sparse conditions through transform switching. In Outdoor (Low Brightness), DWT selection improves temporal alignment with EMD 1.063 versus 1.078 for DTFT, showcasing EECVS's ability to optimize for specific scene characteristics. CES occasionally achieves lower MSE in dynamic scenes but sacrifices temporal fidelity, as evidenced by substantially higher EMD values. TORE exhibits consistently higher errors across all metrics, confirming EECVS's superior compression effectiveness.

These results validate the framework's density-driven adaptive selection strategy: EECVS defaults to DTFT for balanced accuracy across most scenarios, automatically switches to DWT when detecting sparse event patterns, and triggers DCT mode for maximum compression efficiency in dense conditions.

C. Adaptive Coefficient Management

Table I reveals how coefficient budget size affects reconstruction quality across transform types. EMD is essentially unchanged across M for all transforms (≈ 0.964), indicating that coefficient size primarily affects pixel-space fidelity rather than temporal alignment preservation. This finding validates our approach of using fixed temporal windows with variable coefficient budgets.

EECVS's DTFT component benefits significantly from larger coefficient sets. Increasing from $M = 8$ to $M = 24$ reduces MSE by 2.8% while simultaneously improving SSIM from 0.882 to 0.887. DCT demonstrates remarkable stability across budgets variations, with MSE varying minimally (1.731 vs 1.740×10^{-3}). This enables EECVS to deploy DCT efficiently when density demands it. DWT exhibits negligible sensitivity to coefficient count, allowing EECVS to use compact representations for sparse windows without quality degradation.

D. Computational Performance

Table III shows significant computational performance differences across transform types with small coefficient budgets (M=8). the DCT encoder achieves the lowest latency (1.5 ± 0.5 ms) and the highest throughput (2157.3 ± 559.5 kev/s), making it about $2.7\times$ faster than both DTFT (~ 803 kev/s) and DWT ($\sim$ 794 kev/s). The relative efficiency column reflects this gap, with DTFT and DWT operating at roughly 37% of the DCT throughput. While DTFT and DWT exhibit narrower variance in time and speed, DCT still retains substantial headroom even at one standard deviation below its mean performance. DCT still sustains $\gtrsim 1.6$ Mev/s. These measurements validate our scheduling policy: routing high-density or latency-constrained windows to DCT processing while reserving DTFT and DWT for scenarios where their reconstruction characteristics justify additional computational cost.

E. Downstream Task Validation

Downstream evaluation protocol. To isolate the impact of compression on segmentation, we use a compression–decompression pipeline. We first compress the input event stream with each method, then reconstruct (*decompress*) an event stream by applying the corresponding inverse transform using the retained coefficients. The reconstructed events are then passed to EventSAM [25] using the same preprocessing and inference settings across all methods.

EventSAM is not retrained or fine-tuned for any representation. We use the original pretrained weights from [25]. Therefore, performance differences reflect the effect of compression and reconstruction rather than task-specific adaptation.

On EHPT scenes, voxelization achieves the highest per-scene IoU. However, EECVS remains highly competitive. The framework's DCT selection achieves mean IoU 0.78 at 24 channels. This falls within seven points of voxels' 0.85 while providing significantly superior computational efficiency. EECVS matches CES and DTFT performance on EHPT sequences, demonstrating consistent quality preservation.

The framework reveals superior generalization capabilities on MVSEC datasets. While voxelization performance drops catastrophically across all MVSEC scenes, EECVS maintains robust performance through adaptive transform selection. At 24 channels, the framework achieves mean IoU 0.87 compared to voxelization's 0.44 on identical data. This pattern persists across channel configurations, confirming EECVS's adaptive approach provides more generalizable representations than fixed methods.

TABLE II

COMPARISON OF COMPRESSION TECHNIQUES USING SPATIOTEMPORAL METRICS: SPATIAL (MSE↓, SSIM↑) AND TEMPORAL (EMD↓).

	DCT			DTFT			DWT			CES			TORE		
Experiment	**MSE**	**EMD**	**SSIM**	**MSE**	**EMD**	**SSIM**	**MSE**	**EMD**	**SSIM**	**MSE**	**EMD**	**SSIM**	**MSE**	**EMD**	**SSIM**
Indoor (Dark+HDR)	0.0375	1.722	0.832	0.0001	1.102	0.987	0.0646	1.763	0.747	0.0080	1.721	0.977	0.2302	–	0.623
Indoor (Dark+ Blur)	0.0039	1.238	0.903	0.0001	0.288	0.986	0.0055	1.280	0.883	0.0009	1.235	0.983	0.0326	–	0.848
Indoor (HDR + Blur)	0.0236	4.569	0.901	0.0001	0.642	0.990	0.0435	4.461	0.848	0.0050	4.571	0.986	0.1333	–	0.762
Outdoor (Well Lit + Blur)	0.0191	2.050	0.737	0.0004	0.409	0.948	0.0303	2.125	0.643	0.0042	2.108	0.964	0.1413	–	0.489
Outdoor (Dark + Blur)	0.0216	2.448	0.858	0.0003	0.852	0.967	0.0367	2.449	0.782	0.0046	2.452	0.981	0.1393	–	0.657
Indoor (Drone Flying)	0.0093	0.857	0.841	0.0010	0.814	0.958	0.0171	0.870	0.795	0.0013	0.862	0.960	0.0597	–	0.733
Motorcycle (Fast Motion)	0.0080	3.577	0.858	0.0012	1.359	0.936	0.0147	3.613	0.830	0.0011	3.561	0.966	0.0503	–	0.769
Outdoor (Low Brightness)	0.0041	1.097	0.931	0.0011	1.078	0.967	0.0070	1.063	0.921	0.0006	1.118	0.983	0.0251	–	0.887
Outdoor (Dark)	0.0041	4.864	0.913	0.0011	0.826	0.953	0.0080	4.965	0.894	0.0006	4.958	0.978	0.0269	–	0.859

TABLE III

PERFORMANCE COMPARISON OF EVENT-DRIVEN ENCODERS (M=8)

Method	Time (ms)	Speed (kevents/s)	Relative Efficiency
DCT	1.5±0.5	2157.3±559.5	100.0%
DTFT	3.8±0.3	802.8±59.0	37.2%
DWT	3.8±0.2	794.4±40.0	36.8%

TABLE IV

SCENE-WISE COMPARISON OF COMPRESSION TECHNIQUES (M=16). IOU IS MEASURED ON EVENT SEGMENTATION USING EVENTSAM.

Sequence	**CES**	**DCT**	**DTFT**	**DWT**	**Voxel**
EHPT-XC Indoor HDR	0.732	0.734	0.732	0.714	**0.837**
EHPT-XC In-Grain	0.887	0.884	0.887	0.786	**0.929**
EHPT-XC HDR Blur	**0.867**	0.863	**0.867**	0.812	0.855
EHPT-XC Outdoor	0.742	0.730	0.742	0.701	**0.823**
EHPT-XC Outdoor Low	0.689	0.672	0.689	0.664	**0.897**
MVSEC Indoor	0.892	**0.937**	0.892	0.558	0.283
MVSEC Motorcycle	0.672	**0.721**	0.672	0.296	0.381
MVSEC Day1	0.986	**0.992**	0.986	0.939	0.743
MVSEC Night	0.834	**0.839**	0.834	0.575	0.357

EECVS demonstrates remarkable consistency across channel configurations. The framework's overall mean IoU remains stable (0.82 at both 16 and 24 channels), while voxelization exhibits dataset sensitivity regardless of channel count. These results confirm that EECVS provides balanced, generalizable representations through intelligent adaptive selection.

F. Qualitative Assessment

Figures 6- 5 demonstrate EECVS's reconstruction quality across diverse scenarios. The framework's outputs maintain visual similarity to original data at standard viewing distances, with differences concentrated around high-contrast edges and rapid motion regions. DTFT selection preserves

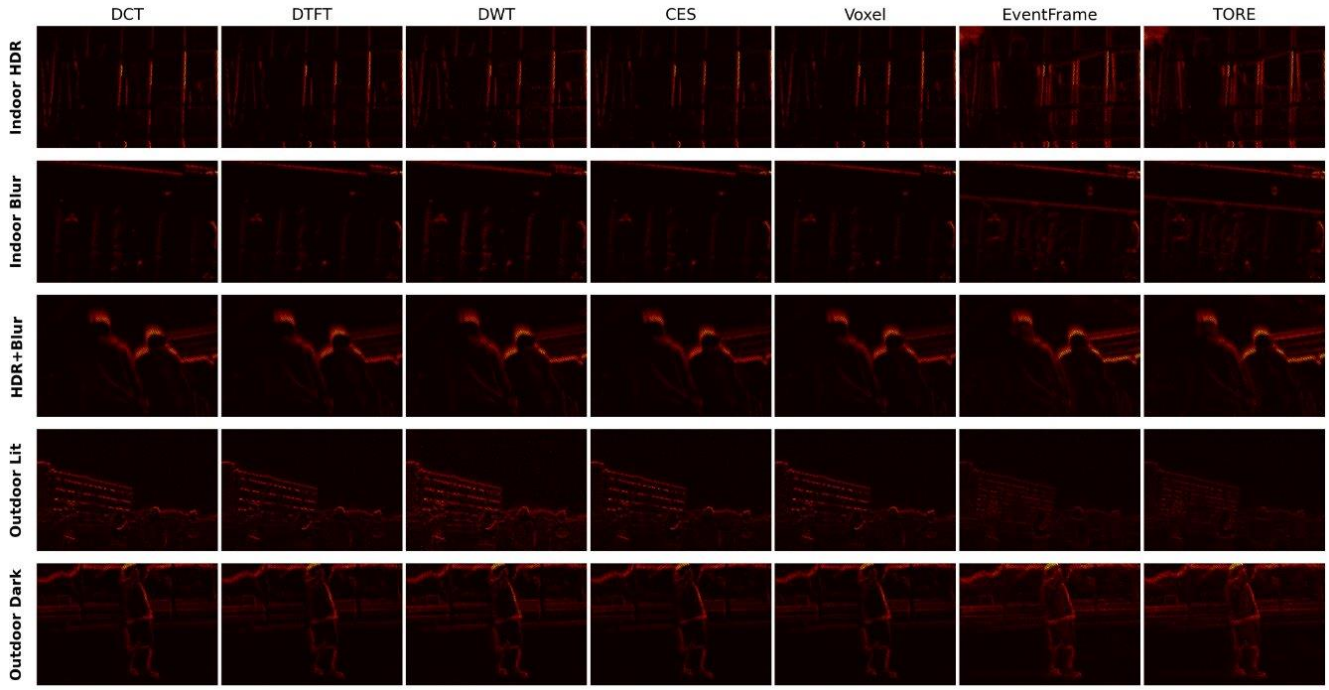

Fig. 5. Qualitative results on the EHPT-XC dataset, showing robustness in high-density, real-world scenarios.

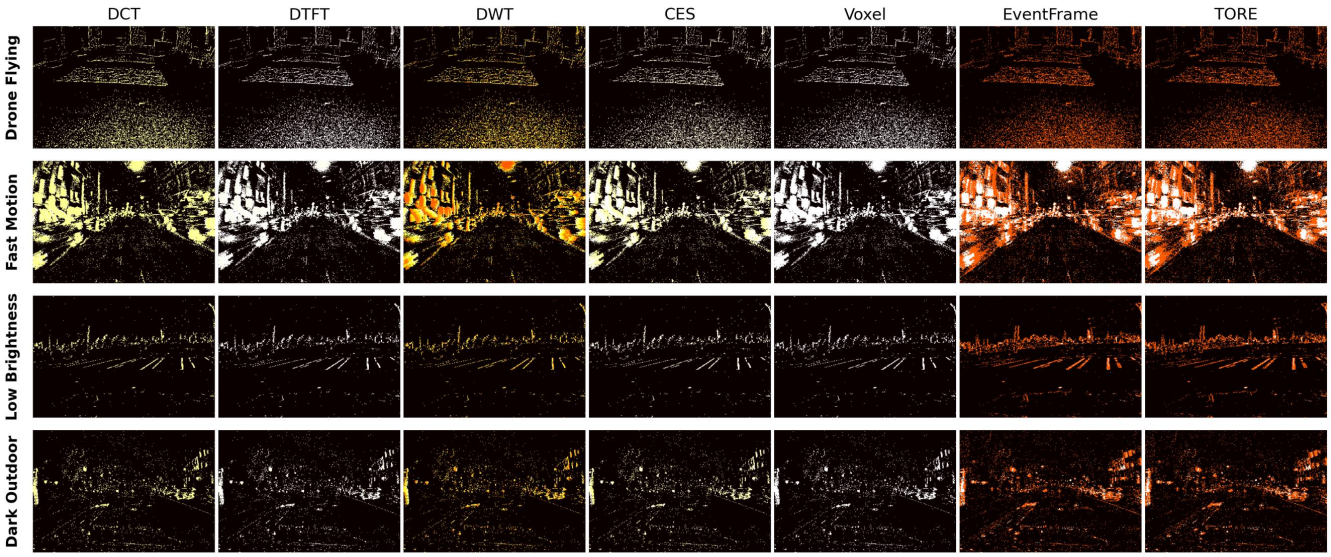

Fig. 6. Qualitative results on the MVSEC dataset, demonstrating preservation of temporal fidelity in dynamic scenes.

TABLE V
SCENE-WISE COMPARISON OF COMPRESSION TECHNIQUES (M=24). IOU IS MEASURED ON EVENT SEGMENTATION USING EVENTSAM.

Sequence	CES	DCT	DTFT	DWT	Voxel
EHPT-XC Indoor HDR	0.739	0.733	0.739	0.706	**0.766**
EHPT-XC In-Grain	0.884	0.884	0.884	0.785	**0.934**
EHPT-XC HDR Blur	**0.867**	0.863	**0.867**	0.809	0.858
EHPT-XC Outdoor	0.733	0.730	0.733	0.689	**0.855**
EHPT-XC Outdoor Low	0.688	0.678	0.688	0.658	**0.839**
MVSEC Indoor	**0.932**	**0.932**	0.925	0.560	0.381
MVSEC Motorcycle	**0.717**	0.710	**0.717**	0.294	0.300
MVSEC Day1	**0.987**	**0.987**	**0.987**	0.940	0.719
MVSEC Night	**0.842**	0.840	**0.842**	0.573	0.364

edge continuity and phase alignment effectively, while DWT mode maintains isolated contours in sparse windows. DCT occasionally introduces minimal ringing artifacts but provides superior computational efficiency. Overall visual distinctions remain modest, validating the framework's adaptive compression approach while maintaining perceptual quality.

V. CONCLUSION

We presented EECVS, the first adaptive event compression framework that fundamentally advances event camera integration in robotics through three core technical innovations. Our continuous-time Dirac impulse model eliminates temporal binning artifacts inherent in existing approaches, enabling precise transform evaluation at exact event timestamps. The density-driven adaptive selection mechanism intelligently chooses among DCT, DTFT, and DWT based on real-time event characteristics, while transform-specific coefficient pruning strategies optimize compression for each domain's sparsity properties. Together, these innovations produce dense representations that preserve both temporal precision and spatial detail while maintaining computational efficiency.

Our evaluation demonstrates EECVS's superior adaptive capabilities: DTFT achieved the lowest earth mover distance in eight scenarios while maintaining excellent MSE and SSIM performance. DCT reached 1.5 millisecond latency with approximately 2.7 times the throughput of alternative transforms at eight coefficients. On cross-dataset evaluation, the framework achieved mean IoU 0.87 at 24 channels on MVSEC while voxel representations managed only 0.44, demonstrating superior generalization capabilities. On EHPT-XC, performance remained competitive within seven points of voxels while providing computational advantages.

This work establishes adaptive compression as a critical capability for robust event camera deployment in robotics. By intelligently matching compression strategies to event stream characteristics, EECVS bridges the gap between event cameras' unique sensing capabilities and standard perception pipelines' dense input requirements. Future extensions will incorporate learned compression strategies, adaptive temporal windowing, and batch optimization for enhanced performance across even broader robotic applications.

REFERENCES

[1] S. Wen, X. Hu, J. Ma, F. Sun, and B. Fang, "Autonomous robot navigation using Retinex algorithm for multiscale image adaptability in low-light environment," *Intelligent Service Robotics*, vol. 12, no. 4, pp. 359–369, Oct. 2019.

[2] Y. Xiao, A. Jiang, J. Ye, and M.-W. Wang, "Making of Night Vision: Object Detection Under Low-Illumination," *IEEE Access*, vol. 8, pp. 123 075–123 086, 2020.

[3] G. Gallego, T. Delbruck, G. Orchard, C. Bartolozzi, B. Taba, A. Censi, S. Leutenegger, A. Davison, J. Conradt, K. Daniilidis, and D. Scaramuzza, "Event-based Vision: A Survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 1, pp. 154–180, Jan. 2022, arXiv:1904.08405 [cs].

[4] M. Gehrig and D. Scaramuzza, "Recurrent Vision Transformers for Object Detection With Event Cameras," 2023, pp. 13 884–13 893.

[5] R. W. Baldwin, R. Liu, M. Almatrafi, V. Asari, and K. Hirakawa, "Time-Ordered Recent Event (TORE) Volumes for Event Cameras," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 2, pp. 2519–2532, Feb. 2023, conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence.

[6] S. Lin, Y. Ma, J. Chen, and B. Wen, "Compressed Event Sensing (CES) Volumes for Event Cameras," *International Journal of Computer Vision*, vol. 133, no. 1, pp. 435–455, Jan. 2025.
[7] X. Lagorce, G. Orchard, F. Galluppi, B. E. Shi, and R. B. Benosman, "HOTS: A Hierarchy of Event-Based Time-Surfaces for Pattern Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 7, pp. 1346–1359, Jul. 2017.
[8] G. Orchard, A. Jayawant, G. K. Cohen, and N. Thakor, "Converting Static Image Datasets to Spiking Neuromorphic Datasets Using Saccades," *Frontiers in Neuroscience*, vol. 9, Nov. 2015, publisher: Frontiers.
[9] B. Zhao, R. Ding, S. Chen, B. Linares-Barranco, and H. Tang, "Feedforward Categorization on AER Motion Events Using Cortex-Like Features in a Spiking Neural Network," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 26, no. 9, pp. 1963–1978, Sep. 2015.
[10] L. Zhu, X. Wang, Y. Chang, J. Li, T. Huang, and Y. Tian, "Event-based Video Reconstruction via Potential-assisted Spiking Neural Network," Mar. 2022, arXiv:2201.10943 [cs].
[11] A. Z. Zhu, L. Yuan, K. Chaney, and K. Daniilidis, "EV-FlowNet: Self-Supervised Optical Flow Estimation for Event-based Cameras," in *Robotics: Science and Systems XIV*, Jun. 2018, arXiv:1802.06898 [cs].
[12] H. Rebecq, R. Ranftl, V. Koltun, and D. Scaramuzza, "Events-To-Video: Bringing Modern Computer Vision to Event Cameras," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2019, pp. 3852–3861, iSSN: 2575-7075.
[13] D. Gehrig, A. Loquercio, K. G. Derpanis, and D. Scaramuzza, "End-to-End Learning of Representations for Asynchronous Event-Based Data," Aug. 2019, arXiv:1904.08245 [cs].
[14] N. Zubić, D. Gehrig, M. Gehrig, and D. Scaramuzza, "From Chaos Comes Order: Ordering Event Representations for Object Recognition and Detection," Aug. 2023, arXiv:2304.13455 [cs].
[15] Y. Nam, M. Mostafavi, K.-J. Yoon, and J. Choi, "Stereo Depth From Events Cameras: Concentrate and Focus on the Future," 2022, pp. 6114–6123.
[16] H. Lou, M. Teng, Y. Yang, and B. Shi, "All-in-Focus Imaging From Event Focal Stack," 2023, pp. 17 366–17 375.
[17] A. Sabater, L. Montesano, and A. C. Murillo, "Event Transformer. A sparse-aware solution for efficient event data processing," Apr. 2022, arXiv:2204.03355 [cs].
[18] S. Schaefer, D. Gehrig, and D. Scaramuzza, "AEGNN: Asynchronous Event-based Graph Neural Networks," Nov. 2022, arXiv:2203.17149 [cs].
[19] B. Xie, Y. Deng, Z. Shao, Q. Xu, and Y. Li, "Event Voxel Set Transformer for Spatiotemporal Representation Learning on Event Streams," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 34, no. 12, pp. 13 427–13 440, Dec. 2024, arXiv:2303.03856 [cs].
[20] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magnetic Resonance in Medicine*, vol. 58, no. 6, pp. 1182–1195, Dec. 2007.
[21] A. B. Konovalov, "Compressed-sensing-inspired reconstruction algorithms in low-dose computed tomography: A review," *Physica Medica*, vol. 124, p. 104491, Aug. 2024.
[22] F. J. Herrmann, M. P. Friedlander, and O. Yilmaz, "Fighting the Curse of Dimensionality: Compressive Sensing in Exploration Seismology," *IEEE Signal Processing Magazine*, vol. 29, no. 3, pp. 88–100, May 2012.
[23] H. Cho, T. Kim, Y. Jeong, and K.-J. Yoon, "A Benchmark Dataset for Event-Guided Human Pose Estimation and Tracking in Extreme Conditions," *Advances in Neural Information Processing Systems*, vol. 37, pp. 134 826–134 840, Dec. 2024.
[24] A. Z. Zhu, D. Thakur, T. Ozaslan, B. Pfrommer, V. Kumar, and K. Daniilidis, "The Multi Vehicle Stereo Event Camera Dataset: An Event Camera Dataset for 3D Perception," *IEEE Robotics and Automation Letters*, vol. 3, no. 3, pp. 2032–2039, Jul. 2018, arXiv:1801.10202 [cs].
[25] Z. Chen, Z. Zhu, Y. Zhang, J. Hou, G. Shi, and J. Wu, "Segment Any Event Streams via Weighted Adaptation of Pivotal Tokens," 2024, pp. 3890–3900.