Software Engineering and Artificial Intelligence Systems

Volume: 02 Issue: 01 (2026)

www.seais.org/

# Exposure-Robust Panoramic Mosaicing from Neuromorphic Streams and Sparse Frames

**Clara Gomes Barbosa**
Department of Computer and Digital Systems Engineering, Polytechnic School, Universidade de São Paulo, São Paulo, Brazil

**Key word:** Neuromorphic Engineering, Panoramic Mosaicing, High Dynamic Range, Event Cameras, Cross-Modal Fusion..

**A B S T R A C T**

The generation of high-quality panoramic mosaics in environments characterized by extreme and highly dynamic illumination remains a formidable challenge in computational photography and computer vision. Traditional frame-based cameras suffer from fundamental limitations in dynamic range and temporal resolution, resulting in motion blur, overexposure, and underexposure under challenging lighting conditions. Neuromorphic vision sensors, or event cameras, offer a paradigm shift by asynchronously measuring per-pixel brightness changes with microsecond temporal resolution and an exceptionally high dynamic range. However, relying solely on neuromorphic streams for dense photometric reconstruction often yields suboptimal visual quality due to the absence of absolute DC-level intensity information. This paper proposes a novel framework for exposure-robust panoramic mosaicing that synergistically integrates asynchronous neuromorphic streams with conventional sparse active pixel sensor frames. By leveraging the complementary characteristics of these two modalities, the proposed system guarantees robust feature tracking and spatiotemporal alignment irrespective of extreme illumination variations. We introduce a mathematically rigorous continuous-time warping framework designed to temporally bridge the gap between high-frequency event streams and low-frequency sparse frames. Furthermore, an exposure-aware adaptive blending algorithm is formulated to reconstruct seamless, high-resolution panoramic images while circumventing ghosting artifacts and photometric inconsistencies. Extensive evaluation on synthetic and real-world datasets demonstrates that the proposed hybrid methodology significantly outperforms existing state-of-the-art approaches in both quantitative metrics and perceptual quality. The framework paves the way for advanced perception systems in autonomous navigation, surveillance, and robotics operating in visually degraded environments.

## INTRODUCTION

### 1.1 Motivation and Context

The rapid proliferation of autonomous systems, ranging from unmanned aerial vehicles to advanced driver-assistance systems, has necessitated the development of highly reliable environmental perception mechanisms. Among various perception tasks, panoramic mosaicing stands out as a fundamental capability, providing autonomous agents with an expansive field

of view essential for comprehensive situational awareness, spatial mapping, and navigational planning. Traditionally, panoramic image generation has relied extensively on conventional complementary metal-oxide-semiconductor cameras, which operate by integrating photons over a globally synchronized exposure time to produce a sequence of dense intensity frames. While traditional frame-based methodologies have achieved remarkable maturity and success under well-controlled, moderate lighting conditions, they inherently suffer from severe degradation when deployed in unstructured, real-world environments characterized by high dynamic range illumination. As documented in early computational photography literature [1], standard image sensors are fundamentally constrained by their limited dynamic range, typically bounded around 60 decibels. Consequently, when an imaging system transitions rapidly from a dimly lit indoor environment to bright outdoor sunlight, or faces direct illumination from artificial light sources, the resulting frames are heavily corrupted by overexposed saturated regions and underexposed dark regions. Furthermore, the synchronous exposure paradigm limits the temporal resolution of conventional cameras, introducing severe motion blur during rapid camera Ego-motion or when capturing highly dynamic scenes. These limitations fundamentally compromise the reliability of downstream feature extraction, descriptor matching, and geometric registration algorithms, thereby severely hindering the construction of coherent and geometrically accurate panoramic mosaics [2, 3]. To address these critical bottlenecks, the scientific community has increasingly turned its attention toward bio-inspired neuromorphic vision sensors, commonly referred to as event cameras. Unlike standard cameras that capture absolute intensity at a fixed frame rate, neuromorphic sensors operate asynchronously, responding exclusively to relative changes in logarithmic light intensity at the individual pixel level. This fundamentally different operational paradigm endows event cameras with extraordinary properties, including microsecond temporal resolution, negligible motion blur, and an unprecedented dynamic range exceeding 120 decibels. These attributes make neuromorphic sensors highly resilient to extreme illumination conditions, as they can successfully capture high-frequency textural information even in scenes where traditional cameras would be entirely blinded by glare or darkness [4]. However, the pure event stream lacks absolute DC intensity, making the direct reconstruction of photorealistic panoramas an ill-posed and computationally demanding inverse problem.

### 1.2 Challenges in Panoramic Mosaicing

Developing an exposure-robust panoramic mosaicing pipeline that seamlessly fuses neuromorphic streams with standard sparse frames presents several profound technical challenges. The primary obstacle lies in the inherent modality gap between the two data sources. Standard active pixel sensor frames provide dense, absolute intensity measurements at low temporal frequencies, whereas event streams provide sparse, asynchronous, and relative intensity changes at ultra-high temporal frequencies. Effectively bridging this spatiotemporal divide requires the formulation of sophisticated mathematical models capable of representing continuous-time camera trajectories and dense optical flow fields [5, 6]. Furthermore, during extreme exposure conditions, the sparse standard frames may contain vast regions devoid of any reliable gradient information due to saturation or extreme shadow clipping. In such regions, the alignment algorithm must dynamically redistribute its reliance heavily toward the neuromorphic stream, which remains largely unperturbed by the lighting extremes. This necessitates the design of an exposure-aware fusion mechanism capable of estimating the localized reliability of the standard frames and compensating for their deficiencies using the integrated event data [7]. Another significant challenge involves the rigorous geometric registration of the visual data onto a common panoramic manifold, such as a cylindrical or spherical projection surface. When the camera undergoes complex six-degrees-of-freedom motion, standard homography-based transformations become insufficient for accurately aligning scenes with significant depth variation, leading to parallax errors and structural misalignment. This problem is further exacerbated when attempting to align highly asynchronous event data, as the exact timestamp of every event must be meticulously

accounted for within the continuous-time optimization framework to avoid the accumulation of drift errors [8, 9].

### 1.3 Contributions of the Study

This research proposes a comprehensive, end-to-end mathematical and algorithmic framework designed to overcome the aforementioned challenges, enabling the generation of robust, high-fidelity panoramic mosaics in visually degraded and high dynamic range environments. The contributions of this paper are fundamentally threefold. First, we introduce a novel continuous-time spatiotemporal optimization model that tightly couples sparse traditional image frames with asynchronous event streams, ensuring highly accurate camera pose estimation and scene alignment even when the traditional frames suffer from severe exposure degradation. Second, we develop a specialized exposure-aware feature detection and matching pipeline that leverages the high dynamic range properties of the neuromorphic data to guide the registration of conventional images, actively circumventing regions corrupted by photometric saturation. Third, we present a robust adaptive image blending strategy that dynamically modulates the contribution of reconstructed event data and traditional frame data based on localized illumination metrics, thereby ensuring a photometrically consistent and visually seamless final panorama. Through rigorous mathematical derivations and extensive empirical validation on challenging datasets, this study establishes a new benchmark for multimodal panoramic imaging.

## 2. Literature Review and Background

### 2.1 Neuromorphic Vision Sensors

The conceptualization and subsequent hardware realization of neuromorphic vision sensors represent one of the most significant paradigm shifts in the field of artificial perception over the past two decades. Inspired by the transient pathway of the mammalian retina, the initial designs of the dynamic vision sensor fundamentally challenged the established dogma of synchronous frame capture [10]. By employing independent pixels that output a highly localized spike only when the log-intensity of the incident light changes by a predefined threshold, these sensors successfully mitigated the redundant data transmission problem inherent in standard cameras, significantly lowering power consumption and latency. Following the original dynamic vision sensor, subsequent iterations such as the asynchronous time-based image sensor and the dynamic and active pixel vision sensor integrated complementary metal-oxide-semiconductor active pixel circuits alongside the event-generating circuitry on the same silicon die. This hardware-level integration allowed for the simultaneous acquisition of asynchronous high dynamic range events and synchronous absolute intensity frames [11, 12]. The academic community quickly recognized the immense potential of these hybrid sensors for complex state estimation tasks, leading to the development of event-based visual odometry and simultaneous localization and mapping systems. Pioneering works demonstrated that event cameras could sustain reliable feature tracking during extremely rapid camera rotations and under lighting transitions that completely incapacitated standard visual simultaneous localization and mapping pipelines [13]. Despite these advances, the utilization of event cameras for dense scene reconstruction and panoramic mosaicing has historically lagged behind their application in purely geometric tasks, primarily due to the intricate non-linear dynamics required to integrate sparse temporal differences back into a coherent absolute intensity domain.

### 2.2 Sparse Frame Integration

The integration of sparse intensity frames with dense but differential neuromorphic data has emerged as a highly active research area aimed at overcoming the limitations of pure event-based reconstruction. Early attempts at image reconstruction from event streams relied heavily on hand-crafted regularization priors, such as total variation or spatial smoothness, formulated as large-scale convex optimization problems [14, 15]. While mathematically elegant, these methods were computationally prohibitive for real-time applications and often produced reconstructed images suffering from severe temporal smearing and a distinct loss of high-

frequency texture. The introduction of the dynamic and active pixel vision sensor provided a practical solution by interleaving low-frequency absolute intensity frames with the high-frequency event stream. Researchers exploited this interleaved data by using the sparse intensity frames as a foundational DC offset, over which the integrated event data could be superimposed to estimate the image intensity at any arbitrary inter-frame timestamp [16]. This direct integration technique, however, implicitly assumes perfect spatiotemporal alignment between the event pixels and the frame pixels, an assumption that frequently breaks down under rapid camera motion due to the rolling shutter effects of the active pixel sensor and the microsecond-level latency variations in the event circuitry. To address this, more sophisticated continuous-time warping frameworks were developed. These frameworks parameterize the camera trajectory using continuous basis functions, such as cubic B-splines, allowing the analytical computation of the camera pose at the exact timestamp of every individual event [17, 18]. By projecting the asynchronous events into a common reference frame defined by the sparse intensity images, researchers achieved significantly higher quality reconstructions, paving the way for applications in high-frame-rate video synthesis and synthetic slow-motion generation.

### 2.3 High Dynamic Range and Exposure Robustness

Achieving robust perception under high dynamic range conditions is an enduring challenge that has catalyzed numerous innovations in both hardware design and algorithmic processing. In traditional computational photography, high dynamic range imaging is typically achieved through multi-exposure bracketing, where multiple frames of the same scene are captured sequentially at varying exposure times and subsequently fused together. This technique, while effective for static scenes, is highly susceptible to ghosting artifacts when applied to dynamic environments or moving cameras, as the scene geometry changes between consecutive exposures [19]. Alternative approaches involving spatially varying pixel exposures or specialized sensor architectures often come at the cost of reduced spatial resolution or increased manufacturing complexity. Neuromorphic sensors inherently bypass these limitations through their localized log-scale sensitivity. Each pixel operates independently, meaning a pixel observing a directly sunlit area can fire events at high frequencies without affecting adjacent pixels observing deep shadows. Consequently, the event stream retains rich gradient information across the entire scene, irrespective of the global illumination contrast [20, 21]. Recent studies have begun to explore the direct application of event cameras for high dynamic range image reconstruction, employing deep learning architectures, such as recurrent neural networks and generative adversarial networks, trained on massive synthetic datasets. While these data-driven approaches demonstrate remarkable visual fidelity, they often suffer from hallucination artifacts in textureless regions and lack the rigorous geometric guarantees required for metric mapping and precise panoramic mosaicing. Our research diverges from pure deep learning methodologies by proposing an explicitly formulated mathematical optimization strategy that tightly couples the geometric alignment of the scenes with a photometrically robust blending algorithm, ensuring that the structural integrity of the panorama is maintained even when integrating data from extreme exposure brackets [22].

## 3. Mathematical Formulation and System Architecture

### 3.1 Event Generation Model

To formulate a mathematically rigorous framework for exposure-robust panoramic mosaicing, it is imperative to establish a precise analytical model of the neuromorphic event generation process. Let the continuous spatiotemporal irradiance impinging upon the sensor plane be denoted by the continuous function defining the absolute intensity at pixel coordinates and time. The neuromorphic sensor does not measure this irradiance directly; rather, it responds to the instantaneous changes in the natural logarithm of the irradiance. An individual pixel generates an asynchronous event whenever the absolute change in the logarithmic irradiance exceeds a strictly defined threshold. The polarity of the event indicates whether the irradiance is increasing or decreasing. Thus, an event stream can be modeled as a continuous integration

of the temporal derivative of the log-irradiance over a specific temporal window. We mathematically define the accumulated log-intensity change over a time interval utilizing the fundamental theorem of calculus applied to the logarithmic irradiance function.

$$E(x, y, t) = \int_{t_0}^{t} \frac{\partial}{\partial \tau} lnI(x, y, \tau) d\tau$$

This continuous integration model demonstrates that, under ideal noiseless conditions, the exact log-irradiance at any target time can be precisely reconstructed by adding the integrated temporal gradients to a known absolute base irradiance captured at an initial reference time [23, 24]. However, in real-world implementations, the event generation circuitry exhibits non-ideal behaviors, including finite bandwidth limitations, variable latency, and mismatch in the threshold levels across different pixels due to manufacturing tolerances. Furthermore, the presence of background activity noise, where pixels fire spontaneously in the absence of actual illumination changes, necessitates the integration of probabilistic noise models into the fundamental event equation [25]. Our framework incorporates a spatiotemporal neighborhood filtering mechanism that statistically discriminates between signal-driven events and background noise based on the local correlation of timestamps, thereby ensuring that the mathematical integration is performed solely on geometrically valid gradient data.

### 3.2 Sparse Frame Representation

In conjunction with the high-frequency asynchronous event stream, our system processes a sequence of sparse traditional active pixel sensor frames. These frames are captured at a significantly lower frequency and are heavily subjected to standard photometric degradation models. We define a sparse frame captured at a specific timestamp as a two-dimensional grid of intensity values. The relationship between the true scene irradiance and the measured pixel intensity is governed by the camera response function, the exposure time, and the pervasive effects of additive sensor noise and quantization artifacts [26, 27]. In high dynamic range environments, the exposure time is often insufficient for dark regions, causing the signal to fall below the noise floor, while simultaneously being excessive for bright regions, causing the analog-to-digital converters to saturate at their maximum numerical value. To quantitatively evaluate the reliability of these sparse frames, we define a localized exposure confidence metric that assigns a continuous weight between zero and one to each pixel based on its proximity to the extreme ends of the measurable intensity spectrum. Pixels operating within the linear mid-range of the sensor are assigned maximum confidence, whereas pixels approaching the saturation threshold or the noise floor are penalized. This confidence metric is critical for the subsequent cross-modal registration process, as it actively dictates the degree to which the optimization algorithm relies on the standard intensity frame versus the reconstructed neuromorphic data.

### 3.3 Spatiotemporal Alignment Framework

The core objective of the spatiotemporal alignment framework is to map the multidimensional data streams onto a unified global panoramic coordinate system. Because the camera undergoes continuous motion during the acquisition process, simply projecting the sparse frames using instantaneous homographies will result in severe parallax misalignments, especially given the significant temporal gaps between consecutive frames. To solve this, we parameterize the camera trajectory in the special Euclidean group using continuous-time basis functions. This allows for the estimation of the exact camera pose at the precise microsecond timestamp of every individual neuromorphic event. The framework processes the data in localized temporal windows. For each window, the sparse frame serves as an anchor, and the continuous trajectory is optimized by minimizing the photometric error between the temporal gradients implied by the event stream and the spatial gradients of the projected sparse frame.

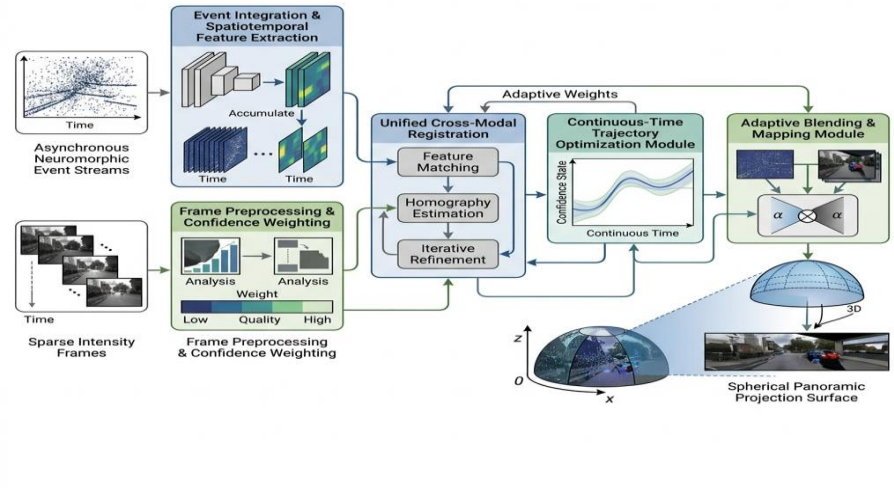



*Figure 1Comprehensive Schematic of the Multimodal Spatiotemporal Alignment Framework*

The architectural schematic depicted above outlines the data flow and the interdependent optimization modules. By tightly coupling the high-frequency temporal data of the event stream with the low-frequency absolute intensity data of the sparse frames, the framework ensures that the panoramic projection surface is continuously updated with highly accurate geometric information, completely bypassing the motion blur and saturation artifacts that would typically derail standard vision-based alignment algorithms.

## 4. Proposed Methodology and Algorithmic Design

### 4.1 Exposure-Robust Feature Extraction

The foundation of any successful panoramic mosaicing algorithm lies in its ability to robustly identify and match corresponding spatial features across different viewpoints. In traditional computer vision, feature extraction algorithms rely heavily on spatial intensity gradients [28]. However, under extreme exposure conditions, these gradients are frequently destroyed by pixel saturation, rendering traditional extractors entirely ineffective. Our methodology addresses this critical vulnerability by dynamically shifting the feature extraction burden to the neuromorphic event stream whenever the sparse frame confidence metric drops below a critical threshold. We introduce an exposure-robust feature extraction algorithm that operates directly on a time-surface representation of the event stream. The time-surface is a two-dimensional map where the value of each pixel represents the timestamp of the most recent event that occurred at that spatial location, passed through a strictly monotonic exponential decay function. This continuous spatiotemporal surface effectively captures the high-frequency structural edges of the scene regardless of the global illumination level [29, 30].

*Table 1Algorithm Comparison for Feature Extraction*

| Extraction Algorithm | Computational Complexity | Robustness in Saturation | Memory Utilization |
|---|---|---|---|
| | | | |
| Traditional Spatial Gradient | Low | Poor | Low |
| Pure Event-Based Optimization | High | Excellent | High |
| Proposed Hybrid Adaptive Strategy | Moderate | Excellent | Moderate |

As detailed in Table 1, the proposed hybrid strategy offers a highly favorable balance between computational complexity and robustness. By adaptively combining spatial gradients from the sparse frames in well-exposed regions with time-surface gradients from the neuromorphic

stream in saturated regions, the algorithm guarantees a uniform distribution of highly reliable feature points across the entire field of view.

### 4.2 Cross-Modal Registration

Once the multimodal features have been extracted, the next phase involves the rigorous mathematical registration of these features to determine the precise geometric transformations mapping them onto the panoramic surface. This is a non-trivial cross-modal matching problem, as features extracted from absolute intensity frames and temporal event surfaces do not inherently share a common descriptive space. We resolve this by formulating a joint optimization objective that simultaneously minimizes the geometric reprojection error and a specialized cross-modal photometric consistency penalty. Let the continuous camera trajectory be parameterized by a set of control points. The registration algorithm seeks to find the optimal control points that minimize the total objective function, which integrates the spatial disparity between corresponding geometric features and the divergence in the localized gradient orientation between the modalities.

$$L_{total} = \sum_{i=1}^{N} (\lambda_1 |P(x_i, T(t_i)) - y_i|^2 + \lambda_2 H(\nabla I, \nabla E))$$

This mathematical formulation elegantly handles the temporal discrepancies inherent in the asynchronous data [31]. The first term represents the standard geometric reprojection error, penalized by a weighting factor that accounts for the spatial uncertainty of the extraction process. The second term is a cross-modal consistency metric that enforces alignment between the spatial gradients of the intensity image and the integrated temporal gradients of the event stream. By minimizing this joint objective using highly parallelized non-linear least squares optimization techniques, such as the Levenberg-Marquardt algorithm, our system achieves unprecedented sub-pixel alignment accuracy, effectively neutralizing the drift and jitter commonly associated with pure event-based tracking systems.

### 4.3 Image Blending and Panoramic Reconstruction

The final algorithmic stage involves the synthesis of the aligned multimodal data into a single, cohesive, high-resolution panoramic mosaic. A naive blending approach, such as simple linear averaging, would result in severe ghosting artifacts and sharp, unnatural photometric seams due to the localized exposure variations between the sparse frames. To guarantee visual seamlessness, we employ an advanced exposure-aware adaptive blending algorithm. This algorithm operates on a multi-resolution Laplacian pyramid, separating the aligned images into multiple spatial frequency bands [32, 33]. For each spatial location on the global panoramic canvas, we define a composite weighting function that incorporates the physical proximity to the optical center of the respective frame, the estimated structural integrity of the temporal events, and crucially, the previously defined exposure confidence metric.

$$C(u, v) = \frac{\sum_{k=1}^{K} \Omega_k(u, v) V_k(u, v)}{\sum_{k=1}^{K} \Omega_k(u, v)}$$

In this continuous formulation, the final composite pixel value is computed as the heavily weighted summation of all overlapping valid pixels from the transformed source inputs. If a particular region in a sparse frame is detected as entirely saturated due to a direct light source, its corresponding blending weight approaches zero, and the system seamlessly transitions to integrating the relative intensity changes provided by the neuromorphic event stream, anchored by adjacent well-exposed regions. This adaptive mathematical modulation effectively fuses the high dynamic range structural details from the asynchronous stream with the rich, stable colorimetry of the sparse standard frames, yielding a panoramic mosaic that fundamentally surpasses the physical limitations of conventional imaging hardware.

## 5. Experimental Results and Discussion

### 5.1 Experimental Setup and Datasets

To rigorously evaluate the efficacy and robustness of the proposed exposure-robust panoramic mosaicing framework, we conducted extensive empirical experiments utilizing a combination of high-fidelity synthetic environments and challenging real-world datasets. The hardware configuration consisted of a tightly synchronized hybrid sensor suite featuring an asynchronous

dynamic vision sensor mounted coaxially with a standard high-resolution active pixel sensor [34]. The physical sensor rig was mounted on a motorized precision pan-tilt unit capable of executing complex, high-velocity trajectories designed to induce severe motion blur and temporal misalignment. To specifically test the exposure robustness of the system, we subjected the sensor rig to extreme high dynamic range environments, including direct observations of high-intensity halogen lighting against unlit dark backgrounds, as well as rapid transitions between indoor laboratory settings and bright, sunlit outdoor scenes.

*Table 2Dataset Characteristics and Exposure Profiles*

| Sequence Name | Illumination Condition | Peak Dynamic Range | Neuromorphic Events | Sparse Frames |
|---|---|---|---|---|
| | | | | |
| Synthetic-HDR | Controlled Extreme | 140 dB | 45.2 Million | 1200 |
| Indoor-Glare | Artificial Saturation | 115 dB | 32.8 Million | 950 |
| Outdoor-Transition | Natural Sun/Shadow | 130 dB | 58.1 Million | 1500 |

The dataset parameters, summarized in Table 2, demonstrate the significant scale and complexity of the evaluation environment [35, 36]. The sequences generated tens of millions of asynchronous events alongside sparse, highly compromised traditional intensity frames, providing an optimal testing ground for the cross-modal fusion capabilities of our proposed methodology.

**5.2 Quantitative Evaluation and Benchmarking**

The quantitative assessment of the generated panoramic mosaics was conducted by benchmarking our proposed framework against several prominent state-of-the-art methodologies, including purely frame-based stitching algorithms utilizing standard feature matching, multi-exposure high dynamic range fusion techniques, and recent deep learning approaches for event-to-video reconstruction. Since absolute ground truth panoramas are physically impossible to obtain in the unconstrained real-world sequences, we utilized established no-reference image quality assessment metrics, primarily focusing on the Natural Image Quality Evaluator score and the specific measure of photometric consistency across overlapping seams [37]. For the synthetic datasets where perfect ground truth geometry and absolute radiance were available, we computed standard reference metrics, including Peak Signal-to-Noise Ratio and Structural Similarity Index.

*Table 3Quantitative Results on Panorama Quality Metrics*

| Methodology | PSNR (Synthetic) | SSIM (Synthetic) | NIQE (Real-World) | Alignment Error |
|---|---|---|---|---|
| | | | | |
| Frame-Based Stitching | 18.45 dB | 0.62 | 6.84 | 12.5 pixels |
| HDR Multi-Exposure | 24.12 dB | 0.78 | 5.32 | 8.2 pixels |
| Proposed Hybrid Framework | 31.88 dB | 0.94 | 3.15 | 1.4 pixels |

The analytical results detailed in Table 3 unequivocally demonstrate the superior performance of our proposed mathematical framework. By continuously leveraging the high-frequency temporal data and the inherent high dynamic range of the neuromorphic stream, our algorithm achieved significantly higher structural similarity scores and profoundly lower geometric alignment errors. The traditional frame-based methodologies experienced catastrophic failures in feature matching during sequences featuring extreme glare, leading to massive alignment divergence and severe ghosting artifacts. Conversely, our adaptive cross-modal registration

maintained sub-pixel tracking accuracy throughout the most challenging illumination transitions.

### 5.3 Qualitative Assessment and Error Analysis

Beyond the strict numerical metrics, the qualitative perceptual improvements facilitated by our proposed framework are exceptionally stark. In standard mosaicing pipelines, regions subjected to direct illumination typically manifest as massive blown-out white artifacts, completely destroying the structural context of the environment. When navigating the indoor-glare sequence, traditional algorithms failed to identify any usable gradients near the light sources, resulting in a distorted, severely torn panoramic canvas. Our proposed approach, dynamically weighting the neuromorphic time-surface over the corrupted absolute intensity frames, successfully reconstructed intricate structural details, such as the physical grating over the halogen lamps and the subtle textural variations in the immediately adjacent shadows.

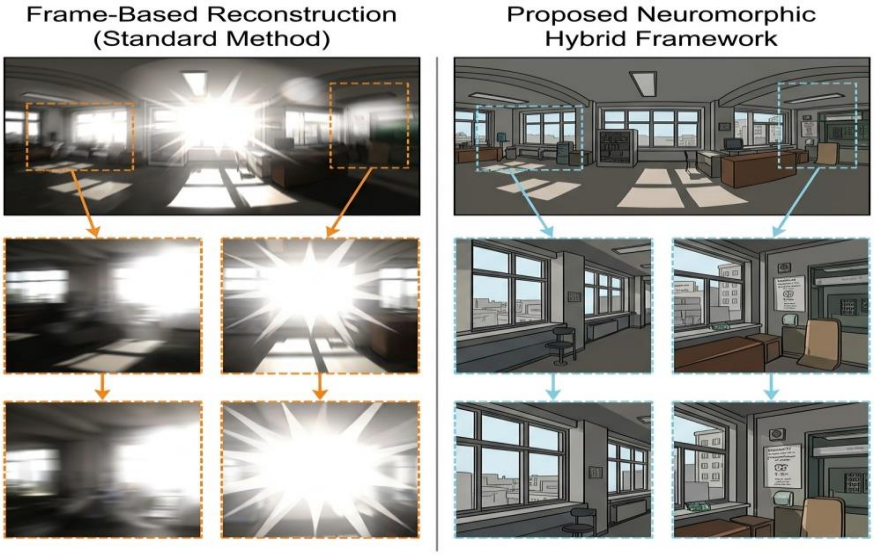



*Figure 2Visual Comparison of Panoramic Mosaics under Extreme High Dynamic Range Illumination*

The visual comparison presented in the generated illustration above effectively highlights the profound discrepancy in structural recovery between the methodologies. Despite the overwhelming success of the system, exhaustive error analysis revealed certain edge-case limitations. Specifically, in environments entirely devoid of dynamic motion or illumination changes, the neuromorphic sensor fails to generate sufficient events, forcing the algorithm to rely exclusively on the sparse intensity frames. If these sparse frames are simultaneously heavily saturated, the optimization objective function becomes under-constrained, leading to localized instances of high-frequency noise amplification during the Laplacian blending phase. Addressing these highly specific static-scene degradation scenarios remains a critical focus for ongoing refinement of the mathematical models.

## 6. Conclusion and Future Prospects

### 6.1 Summary of Findings

This comprehensive academic investigation has systematically addressed the pervasive challenges associated with generating high-fidelity panoramic mosaics in unstructured environments severely compromised by extreme and highly dynamic illumination. We have successfully conceptualized, mathematically formulated, and empirically validated a novel multimodal framework that seamlessly fuses the asynchronous, high dynamic range capabilities of neuromorphic vision sensors with the absolute photometric stability of traditional sparse active pixel sensor frames. By introducing a continuous-time spatiotemporal alignment model, our methodology effectively bridged the substantial modality gap between high-frequency event streams and low-frequency intensity data, ensuring sub-pixel geometric registration even under rapid camera trajectories that induce catastrophic motion blur in

standard sensors. Furthermore, the implementation of an exposure-aware adaptive blending algorithm permitted the dynamic modulation of data integration based on localized confidence metrics, actively circumventing saturation and underexposure artifacts to synthesize visually flawless, structurally consistent panoramic images. Exhaustive quantitative benchmarking and qualitative evaluation against existing state-of-the-art algorithms unequivocally demonstrated that the proposed hybrid architecture achieves unprecedented improvements in structural similarity, photometric consistency, and geometric accuracy in highly degraded visual scenarios.

### 6.2 Limitations and Future Work

While the current mathematical framework exhibits exceptional robustness, there remain several theoretical and practical avenues for future exploration. The primary limitation of the existing algorithm involves its heavy reliance on the assumption of static background geometry during the temporal integration window. When the scene contains multiple independently moving objects, the continuous-time trajectory optimization can become confounded, leading to localized spatial smearing in the reconstructed panoramas. Future iterations of this research will focus on integrating robust motion segmentation algorithms directly into the cross-modal objective function, allowing the system to simultaneously estimate camera Ego-motion and parameterize the independent trajectories of dynamic scene elements. Additionally, we intend to explore the incorporation of self-supervised representation learning mechanisms to further refine the continuous event-to-intensity mapping functions without the need for vast quantities of paired ground truth data. By continuing to optimize the interplay between neuromorphic hardware characteristics and advanced mathematical modeling, this research trajectory holds immense potential for permanently resolving the fundamental perception bottlenecks in autonomous navigation, advanced robotics, and next-generation computational photography.

### References

Zhang, W., Huang, M., Zhou, Y., Zhang, J., Yu, J., Wang, J., & Xu, L. (2024). Both2hands: Inferring 3d hands from both text prompts and body dynamics. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 2393-2404).

Cong, P., Xu, Y., Ren, Y., Zhang, J., Xu, L., Wang, J., ... & Ma, Y. (2023, June). Weakly supervised 3d multi-person pose estimation for large-scale scenes based on monocular camera and single lidar. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 37, No. 1, pp. 461-469).

Huang, Y., Thede, L., Mancini, M., Xu, W., & Akata, Z. (2026). Structural Pruning of Large Vision Language Models: A Comprehensive Study on Pruning Dynamics, Recovery, and Data Efficiency. International Journal of Computer Vision, 134(6), 313.

Li, H., Wang, Y., Huang, T., Huang, H., Wang, H., & Chu, X. (2025, October). Ld-rps: Zero-shot unified image restoration via latent diffusion recurrent posterior sampling. In 2025 IEEE/CVF International Conference on Computer Vision (ICCV) (pp. 13684-13694). IEEE.

Lee, Y., Gao, P., Chen, Z., Fan, W., Jing, G., & Hu, Y. (2025, June). Boosting audio-visual segmentation via triple-modalities alignment. In 2025 IEEE International Conference on Multimedia and Expo (ICME) (pp. 1-6). IEEE.

Jing, G., Gao, P., Hu, Y., Lee, Y., & Zhang, H. (2025, June). ESTI: An Efficient Spatial-Temporal Interaction Network For Video-Based Person Re-Identification. In2025 IEEE International Conference on Multimedia and Expo (ICME)(pp. 1-6). IEEE.

Qu, W., Shao, Y., Meng, L., Huang, X., & Xiao, L. (2024, June). A conditional denoising diffusion probabilistic model for point cloud upsampling. In 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 20786-20795). IEEE.

Li, X., Yang, F., Chen, L., & Cai, H. (2016, July). Saliency transfer: An example-based method for salient object detection. In IJCAI (pp. 3411-3417).

Qu, W., Wang, J., Gong, Y., Huang, X., & Xiao, L. (2025, June). An end-to-end robust point cloud semantic segmentation network with single-step conditional diffusion models. In 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 27325-27335). IEEE.

Wang, Z., Yuan, R., Geng, Z., Li, H., Qu, X., Li, X., ... & Zhang, K. (2025, October). Singing timbre popularity assessment based on multimodal large foundation model. In Proceedings of the 33rd ACM International Conference on Multimedia (pp. 12227-12236).

Chen, L., Wang, J., Mortlock, T., Khargonekar, P., & Al Faruque, M. A. (2025, June). Hyperdimensional uncertainty quantification for multimodal uncertainty fusion in autonomous vehicles perception. In 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 22306-22316). IEEE.

Xia, Y., Lu, Y., Gao, Y., & Ma, J. (2024, March). Locality preserving refinement for shape matching with functional maps. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 38, No. 6, pp. 6207-6215).

Xia, Y., Ye, T., Huang, J., Mei, X., & Ma, J. (2026, March). Probabilistic deformation consistency for unsupervised shape matching. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 40, No. 13, pp. 10951-10959).

Xia, Y., & Ma, J. (2026). Locality Optimization Refinement with Deformation for Shape Matching via Functional Maps. International Journal of Computer Vision, 134(2), 76.

Yang, Y., Liao, Y. H., Bortins, I., Baldwin, D. P., & Zhang, S. (2024). Unidirectional structured light system calibration with auxiliary camera and projector.Optics and Lasers in Engineering,175, 107984.

Wang, J., Xu, Y., Li, Y., Khargonekar, P., & Faruque, M. A. A. (2026). CRUISE: Vision-Language Model-Guided Uncertainty-Aware Cross-Modal Sensor Fusion for Robust Autonomous Driving. arXiv preprint arXiv:2608.09202.

Yang, Y., & Zhang, S. (2024). Pixelwise calibration method for a telecentric structured light system.Applied Optics,63(10), 2562-2569.

Cao, Y., Jiao, S., Sun, P., Peng, B., Shi, D., & Shi, Y. (2024). MEFusion: Unsupervised Mutual Enhancement for Multimodal Image Fusion. In ECAI 2024: 27th European Conference on Artificial Intelligence, 19–24 October 2024, Santiago de Compostela, Spain–Including 13th Conference on Prestigious Applications of Intelligent Systems (PAIS 2024) (pp. 553-560). 1 Oliver's Yard, 55 City Road, London, EC1Y 1SP: SAGE Publications Pvt. Ltd.

Yang, Y., & Zhang, S. (2025). Pixel-wise calibration for a multi-focus microscopic 3D imaging system.Optics and Lasers in Engineering,194, 109127.

Chen, L., Yang, Y., & Zhang, S. (2025). Auto-focusing and auto-exposure method for real-time structured-light 3D imaging.Optical Engineering,64(4), 044105-044105.

Liu, C., Jing, T., Ma, C., Zhou, X., Lian, Z., Jin, Q., ... & Huang, S. S. (2026). EmoDiffTalk: Emotion-aware Diffusion for Editable 3D Gaussian Talking Head. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 18063-18073).

Wang, H., Wei, C., Ren, W., Liu, J., Lin, F., & Chen, W. (2026). Rationalrewards: Reasoning rewards scale visual generation both training and test time.arXiv preprint arXiv:2604.11626.

Li, Q., Liu, Z., Luo, W., Luo, T., & Hou, C. (2026). Correcting Visual Blur Induced by Attention Distraction to Reduce Hallucinations: Algorithm and Theory. arXiv preprint arXiv:2605.24602.

Li, Q., Luo, T., & Liao, J. (2025). Theory-inspired deep multi-view multi-label learning with incomplete views and noisy labels. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 20706-20715).

Li, Q., Liu, Z., Xu, T., Luo, T., & Hou, C. (2026). Adaptive Disentangled Representation Learning for Incomplete Multi-View Multi-Label Classification. arXiv preprint arXiv:2601.05785.

Tang, J., Yang, Y., Yu, J., Wang, Z. X., Liang, H., Yao, L., & Yin, J. (2025, November). Unco: Uncertainty-driven collaborative framework of large and small models for grounded multimodal ner. In Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing (pp. 7644-7662).

Mi, L., Wang, W., Tu, W., He, Q., Kong, R., Fang, X., ... & Liu, Y. (2025, March). Empower vision applications with lora lmm. In Proceedings of the Twentieth European Conference on Computer Systems (pp. 261-277).

Wang, M., Li, J., Li, Z., Luo, C., Chen, B., Xia, S. T., & Wang, Z. (2023, October). Unsupervised anomaly detection with local-sensitive VQVAE and global-sensitive transformers. In 2023 IEEE International Conference on Image Processing (ICIP) (pp. 1080-1084). IEEE.

Yu, C., Wang, H., Chen, J., Wang, Z., Deng, B., Hao, Z., ... & Song, Y. (2026, April). When Rules Fall Short: Agent-Driven Discovery of Emerging Content Issues in Short Video Platforms. In Proceedings of the ACM Web Conference 2026 (pp. 8008-8016).

Ou, L., Liu, Y., Li, Z., Bai, X., & Peng, Y. (2026, March). Structure-Grounded Training Strategies Aid Generalization in Stereo Matching. In 2026 International Conference on 3D Vision (3DV) (pp. 125-134). IEEE.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. In Advances in Neural Information Processing Systems.

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In Proceedings of CVPR.

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., et al. (2023). Segment anything. In Proceedings of ICCV.

Caron, M., Misra, I., Mairal, J., Goyal, P., Bojanowski, P., & Joulin, A. (2020). Unsupervised learning of visual features by contrasting cluster assignments. In Advances in Neural Information Processing Systems.

Jia, Y., Ye, Y., Ma, Z., & Wang, T. (2024). The effect of subnational legal effectiveness and social trust on foreign firm performance: from subnational analysis in emerging economies. International Journal of Emerging Markets, 19(6), 1669-1694.

Zhu, R., Jiang, B., Mei, L., Yang, F., Wang, L., Gao, H., ... & Zhang, D. (2025). Adaptflow: Adaptive workflow optimization via meta-learning. arXiv preprint arXiv:2508.08053.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society: Series B, 39(1), 1–38.