

Fast-Rotating Panoramic HDR Reconstruction from Single-Second Spikes

Yuzhu Tao*
Beijing Jiaotong University
Beijing, China
taoyuzhu@bjtu.edu.cn

Shikui Wei*
Beijing Jiaotong University
Beijing, China
shkwei@bjtu.edu.cn

Lang Nie
Chongqing University of Posts and
Telecommunications
Chongqing, China
nielang@cqupt.edu.cn

Qingwen Li
China Tower Corporation Limited
Beijing, China
liqw3@chinatowercom.cn

Bo Li
China Tower Corporation Limited
Beijing, China
libo88872@chinatowercom.cn

Boxin Shi†
Peking University
Beijing, China
shiboxin@pku.edu.cn

Yao Zhao*
Beijing Jiaotong University
Beijing, China
yzhao@bjtu.edu.cn

Yakun Chang‡*
Beijing Jiaotong University
Beijing, China
ykchang@bjtu.edu.cn

Abstract

Conventional digital cameras suffer from limited shutter speed and dynamic range, making it difficult to capture blur-free and high-dynamic-range (HDR) images in fast-rotating scenes, which complicates panoramic image reconstruction. The high-speed spike cameras offer advantages such as high temporal resolution and high dynamic range, making them suitable for capturing panoramic HDR scenes within a short time. In this paper, we present a new framework for reconstructing panoramic HDR color images from single-second spikes. We first introduce a spike-based panoramic reconstruction method that leverages the differential of spike firing time (DSFT) representation for finer texture preservation and spike-noise suppression, followed by DSFT-enhanced vertical constraint matching and texture optimization for reconstruction. Since spike cameras inherently lack colors, the reconstructed panoramic images remain monochromatic. To incorporate color information, we employ a side-by-side conventional color camera that captures a single color frame instantaneously. This color guidance is subsequently integrated through a fine-tuned image editing model. To train and validate practicality, a comprehensive dataset including synthetic and real-world spikes is presented. Experimental results demonstrate that our method generates high-quality panoramic HDR color images in fast-rotating scenes.

*Institute of Information Science, Beijing Jiaotong University; Visual Intelligence +X International Cooperation Joint Laboratory of the MoE.

†State Key Laboratory of Multimedia Information Processing, School of Computer Science, Peking University; National Engineering Research Center of Visual Technology, School of Computer Science, Peking University; and PKU-AI² Robotics Joint Lab of Embodied AI.

*Corresponding author.



This work is licensed under a Creative Commons Attribution 4.0 International License. *MM '26, Rio de Janeiro, Brazil.*

© 2026 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-2213-4/2026/11

<https://doi.org/10.1145/3767308.3835660>

CCS Concepts

• Computing methodologies → Computational photography.

Keywords

HDR imaging; panoramic image; spike camera.

ACM Reference Format:

Yuzhu Tao, Shikui Wei, Lang Nie, Qingwen Li, Bo Li, Boxin Shi, Yao Zhao, and Yakun Chang. 2026. Fast-Rotating Panoramic HDR Reconstruction from Single-Second Spikes. In *Proceedings of the 34th ACM International Conference on Multimedia (MM '26)*, November 10–14, 2026, Rio de Janeiro, Brazil. ACM, New York, NY, USA, 13 pages. <https://doi.org/10.1145/3767308.3835660>

1 Introduction

Panoramic image reconstruction aims to capture and representing a wide field of view (FOV) in a single seamless image. From traditional geometric transformation models [29, 35, 48] to recent data-driven learning methods [23, 38], panoramic image reconstruction with a single camera is extensively addressed in general low-speed scenes. However, as illustrated in Fig. 1 (a), due to limitations in shutter speed and dynamic range, conventional digital cameras often fail to capture blur-free and high-dynamic-range (HDR) images during fast rotation, making it difficult to reconstruct the desired panoramic image. Although multi-exposure approaches [11, 36] are proposed to address the HDR limitation, the limited frame rate makes them still ineffective in fast-rotating scenes, e.g., 180° per second.

In recent years, spike cameras [5, 16, 19] have demonstrated significant potential for high-speed applications. Unlike conventional frame-based high-speed cameras, spike cameras capture the scenes at sampling rates over 20,000 Hz while reading out data in a compact single-bit representation (shown in Fig. 1 (b)). In addition, to maximize the perception of incoming light, most spike cameras are designed in a monochromatic manner, which provides a high dynamic range (≥ 100 dB) [5], making them highly effective

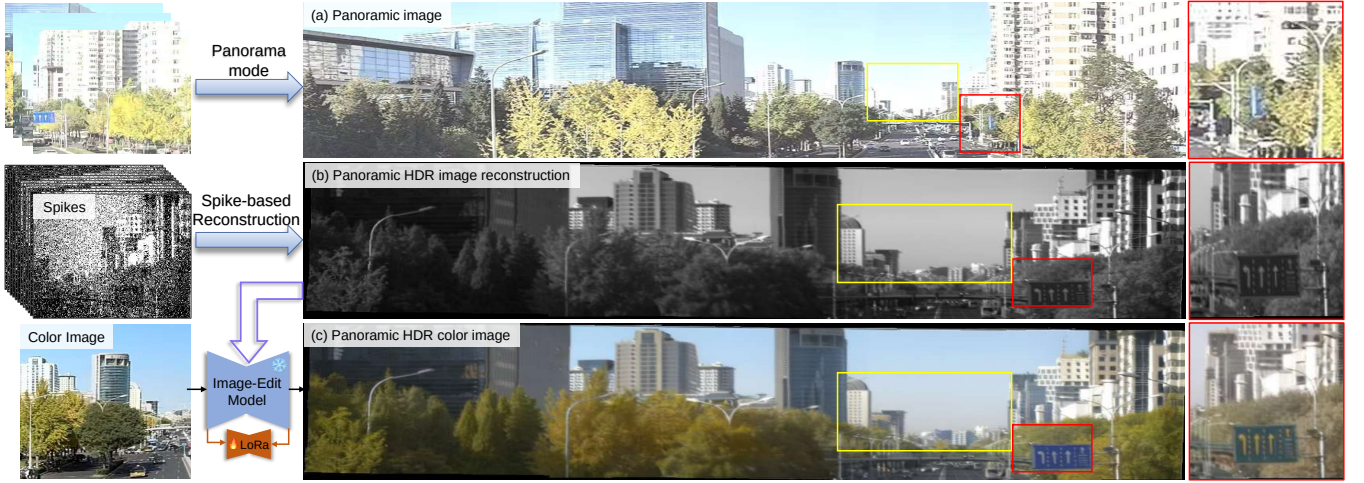


Figure 1: Panoramic HDR color image reconstruction in fast-rotating scenes (180° per second). (a) Conventional digital cameras fail to produce consistent HDR panoramas, resulting in motion blurs (red boxes) and overexposure (yellow boxes). (b) Panoramic HDR monochromatic images from pure spikes. (c) Colorization by finetuning an image-editing model [42] the side-by-side color image as visual prompt.

for recovering fine details that are often lost in conventional color frames [16]. Therefore, with the high temporal resolution and high dynamic range, the spike camera holds significant advantages for image reconstruction in fast-rotating scenes.

However, achieving panoramic HDR reconstruction from spikes is inherently non-trivial and cannot be treated as a direct extrapolation of existing spike-based methods [45, 49, 50, 55]. The fast rotation induces complex spatiotemporal overlaps, where cumulative geometric errors and inherent spike noise lead to misalignment between disparate viewpoints. Furthermore, the sheer volume of reconstructed images necessitated by the high sampling rate imposes a prohibitive computational overhead, creating a significant bottleneck for large-scale panoramic reconstruction. To address these challenges, We first reconstruct images using texture from window (TFW) [55] from spikes. Since TFW may introduce motion blur, we adopt differential spike firing time (DSFT) [53] to provides finer texture details. Consequently, we design a strategy that incorporates DSFT-enhanced vertical constraint matching and texture optimization for reconstruction. Subsequently, we leverage DSFT to provide denser correspondences and vertical geometric constraints for removing misalignments. The aligned images are then seamlessly integrated through linear projection weighting combined with joint bilateral filtering. This approach achieves texture enhancement and facilitate panoramic HDR image reconstruction.

Therefore, as shown in Fig. 1 (b), the temporally dense spikes obtained during fast camera rotation are fully exploited to reconstruct the panoramic HDR image. However, since the spike camera is designed in a monochromatic manner, the output lacks color information. To address this, as shown in Fig. 1 (c), we introduce a conventional color camera positioned side-by-side to capture a single color frame as a visual prompt for panoramic colorization. Unlike the existing hybrid architecture [5] that relies on a beam splitter and spatiotemporal alignment, our simplified hardware setup

eliminates these constraints by leveraging the strong generative priors of large vision-language models. Specifically, we fine-tune an image-editing model [42] to perform panoramic colorization, where the color frame serves as a prompt rather than a pixel-wise reference, thus enabling robust color recovery without spatial alignment. Furthermore, recognizing the lack of training data tailored for our unique configuration, we create a comprehensive synthetic dataset for HDR panoramic reconstruction in fast-rotating scenes. To further validate the practicality in real-world scenes, we also collect real-world data using our side-by-side camera system. Each spike sequence is paired with a high-resolution color image that serves as the visual prompt. Our contributions can be summarized as:

- a side-by-side hybrid camera system for reconstructing panoramic HDR color images without the need of spatial and temporal alignment;
- a spike-based panoramic reconstruction method and a colorization module based on a fine-tuned multimodal large language model for panoramic HDR color image reconstruction; and
- a comprehensive dataset contains synthetic and real-world spikes, supporting HDR panoramic reconstruction, colorization, and fast-rotating scene synthesis.

2 Related work

Panoramic image with conventional cameras. Panoramic image reconstruction with a single camera typically estimates geometric transformations to align sequential images into a seamless panorama. Early methods rely on a global homography [2], while later approaches adopt more flexible local transformations [10, 24, 47] to better handle parallax. Zhang et al. [48] further extend the mesh-based APAP framework [47] by introducing scale-preserving

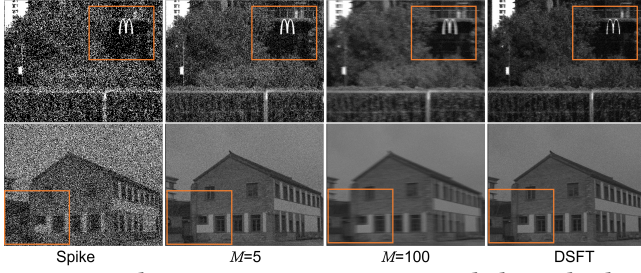


Figure 2: Spike generation. We present single-bit spike data and the reconstructed images produced by the texture from window (TFW) methods ($M = 5$ and $M = 100$) and the DSFT-based method. The orange boxes indicate motion blur.

constraints for more natural panoramic reconstruction. For image blending, weighted averaging [44] is followed by gradient-domain techniques [13, 22] to improve visual consistency. Recent advances in deep learning have introduced supervised [12, 30], weakly supervised [38], and unsupervised [20, 31, 32] paradigms, substantially improving stitching quality. Building upon this trend, Liang et al. [23] extend the parallax-tolerant approach to reconstruct wide-field panoramas. However, experimental evidence [33] and other studies [39] suggest that traditional stitching algorithms, due to the robust priors from feature-based detection, remain more effective for cross-scene and panoramic stitching.

Image reconstruction with spikes. Reconstructing images from spikes typically relies on the imaging model of spike cameras [9, 19, 51, 52]. Early methods reconstruct textures by accumulating spikes within a temporal window or estimating inter-spike intervals [54, 55], but suffer from severe noise due to limited photon counts. To improve reconstruction quality, Chang et al. [4] enhance the signal-to-noise ratio via a rolling mixed-bit acquisition scheme, while Zhao et al. propose Spk2ImgNet [49], a hierarchical network for progressive spike fusion. However, reconstructing HDR images from spikes remains challenging as the single-bit data is insufficient to cover the high dynamic range of real-world scenes. Chang et al. [5] propose a hybrid spike-RGB camera system that performs spatial alignment and frame interpolation simultaneously. However, this hybrid system demands precise synchronization and optical alignment, and the additional space required by the beam splitter complicates the design of compact devices.

Image colorization with guided information. Guided image colorization exploits external cues, including user-provided color hints [43, 46], reference images [1, 17, 40], and text descriptions [27], to produce controllable and semantically consistent results. More recently, large generative models have further advanced language-guided colorization and editing [6, 7, 41, 42]. Unlike these methods, our approach leverages localized color information for panoramic image colorization, preserving global style consistency and realism without explicit spatial correspondence estimation.

3 Preliminaries

In this section, we introduce the spike imaging model and discuss the trade-off between HDR and sharpness.

HDR reconstruction from spikes. Each pixel in a spike camera continuously integrates photogenerated electrons until reaching a

predefined threshold. During integration, while the accumulated charge remains below this threshold, the readout circuit samples a pixel value of 0 at each sampling instant. Once the accumulated electrons exceed the threshold, a spike (encoded as 1) is fired and the charge reservoir is reset to 0. The output of the pixel is

$$S(t) = \begin{cases} 1, & E_c(t) \geq E_{th}, \\ 0, & \text{otherwise,} \end{cases} \quad (1)$$

where $E_c(t)$ and E_{th} denote the accumulated charge and the preset threshold, respectively. An image can be reconstructed by integrating spikes over a time window W_t :

$$I(W_t) = \sum_{i=0}^{M-1} S(t + i\tau), \quad (2)$$

where τ is the sampling interval and $W_t = M \cdot \tau$ is the window length. As demonstrated by Chang et al. [3, 4], increasing M in static scenes helps expand the dynamic range, even when the input data remains in single-bit form. The high temporal density of spike cameras ensures that even a single-second recording provides sufficiently large M for reconstruction. For more general dynamic scenes with rapid camera rotation, we compute dense motion fields from spikes to achieve spatial alignment of textures, thereby enabling the selection of a large M for panoramic reconstruction. Consequently, during camera rotation, all captured spikes are effectively utilized for our task.

Trade-off between HDR and sharpness. Since spike data contains inherent noise, achieving pixel-level precision in spatial alignment remains challenging, which compromises image sharpness when using large values of M . The DSFT [53] overcomes the constraints imposed by the temporal window M and preserves fine texture details. We denote DSFT as $D(t)$, which is computed as follows:

$$D(t) = \min\{T^+(t) - T^-(t), W_{\max}\},^1 \quad (3)$$

where $T^+(t)$ and $T^-(t)$ represent the nearest spike timestamps before and after time t , respectively, and $W_{\max}$ defines the maximum search range to prevent unbounded searching when no spike is triggered. As shown in Fig. 2, using a short integration window (small M) leads to noticeable noise in the reconstructions, while a long integration window (large M) causes temporal aggregation that results in significant motion blur (orange boxes). The DSFT method preserves fine-grained texture details.

4 Method

Pipeline overview. We propose a framework for reconstructing panoramic HDR color images using a monochromatic spike camera and a reference color image as visual guidance. As illustrated in Fig. 3, the captured spike stream is first reconstructed into HDR images and then geometrically aligned to generate a panoramic HDR image through adaptive image registration and texture-aware blending. Finally, an image editing model [42] fine-tuned with LoRA [18] leverages the reference color image as a visual prompt to perform generative colorization without explicit spatial alignment, producing the final panoramic HDR color image.

¹Images generated by the DSFT formula are inverted, so their reciprocals are displayed.

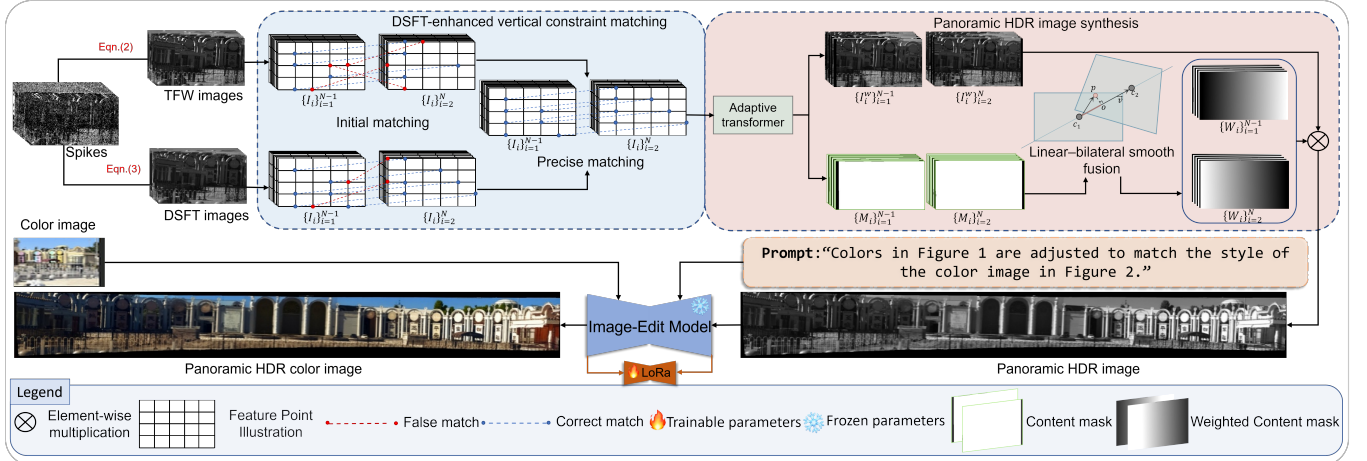


Figure 3: The proposed framework for panoramic HDR color image reconstruction consists of a spike-based panoramic reconstruction method and a fine-tuning image-editing module for colorization. Here, $\{I_i\}_{i=1}^{N-1}$ and $\{I_i\}_{i=2}^N$ denote adjacent image pairs used for matching, where N is the total number of images. $\{I_i^w\}_{i=1}^{N-1}$ and $\{I_i^w\}_{i=2}^N$ denote images aligned to the same coordinate system, $\{M_i\}_{i=1}^{N-1}$ and $\{M_i\}_{i=2}^N$ represent the corresponding masks for the previous and next images, respectively. $\{W_i\}_{i=1}^{N-1}$ and $\{W_i\}_{i=2}^N$ are the associated weighted masks used for reconstruction.

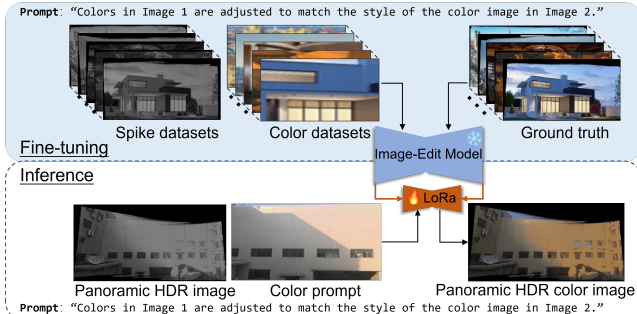


Figure 4: Fine-tuning the Qwen-Image-Edit-2509 [42] with LoRA [18] under a supervised fine-tuning (SFT) paradigm using images reconstructed from spike data as inputs, color datasets and textual prompts as guidance, and initial images as ground truth. During inference, the model generates colorized outputs from monochrome inputs, guided by text and reference color prompt.

4.1 DSFT-enhanced vertical constraint matching

DSFT enhancement. Given the inherent noise in spike cameras, robust feature correspondences are essential for accurate alignment and help mitigate cumulative errors in panorama synthesis. Since images reconstructed via Eqn. (2) may contain motion blur, we further leverage the DSFT representation from Eqn. (3), which exhibits lower sensitivity to motion, to enhance feature matching. By integrating window-based accumulation with DSFT enhancement, our approach combines stable intensity cues with fine-grained textures, thereby improving matching robustness under noisy and motion blur.

Vertical distance constraint. Given the limited inter-frame motion at high frame rates, correspondences exhibiting abnormally large vertical displacement are treated as incorrect matches. As illustrated in Fig. 3, we utilize grid points to simulate feature points

extracted by the SIFT [25] operator and employ the KNN [8] algorithm for feature matching. In this visualization, blue lines indicate correct matches, while red lines represent incorrect matches. Since these mismatches compromise the accuracy of subsequent image alignment, we introduce a vertical distance geometric constraint as a metric for candidate matches. Specifically, let the tentative correspondences be $\mathcal{M} = \{(p_i, q_i)\}_{i=1}^N$, where $p_i = (x_i^p, y_i^p) \in \text{Image}_1$ and $q_i = (x_i^q, y_i^q) \in \text{Image}_2$. For each pair, we compute the vertical residual $d_i = |y_i^p - y_i^q|$ and estimate its empirical mean μ_d and variance σ_d^2 . We retain matches satisfying $d_i \in [T_{\text{low}}, T_{\text{high}}]$, the thresholds and filtered set are

$$\mathcal{M}^* = \{(p_i, q_i) \in \mathcal{M} \mid \mu_d - 2\sigma_d \leq d_i \leq \mu_d + 2\sigma_d\}. \quad (4)$$

This 2σ screening rejects correspondences whose vertical residuals are statistically inconsistent with the inlier distribution. Finally, we employ random sample consensus (RANSAC) [14] to further suppress the effects of noise and ensure reliable correspondences.

4.2 Panoramic HDR image reconstruction

To reconstruct a panoramic HDR image from the recovered HDR frames, we perform adaptive geometric alignment followed by texture-aware image blending.

We first estimate a global transformation followed by mesh-based local refinement using Moving-DLT [47], yielding warped images $\{I_i^w\}_{i=1}^N$ in a unified coordinate system. The corresponding masks $\{M_i\}_{i=1}^N$ are then refined via linear projection weighting and joint bilateral filtering to obtain weighted masks $\{W_i\}_{i=1}^N$, which are used for seamless panoramic HDR reconstruction.

Specifically, we illustrate the image reconstruction process using two adjacent images as an example. First, we obtain the corresponding masks M_1 and M_2 for images I_1^w and I_2^w , respectively. We then compute the geometric centers c_1 and c_2 of their valid regions and construct the connecting direction vector: $v = c_2 - c_1$. For each pixel $p \in M_1 \cap M_2$ in the overlapping region, the normalized projection

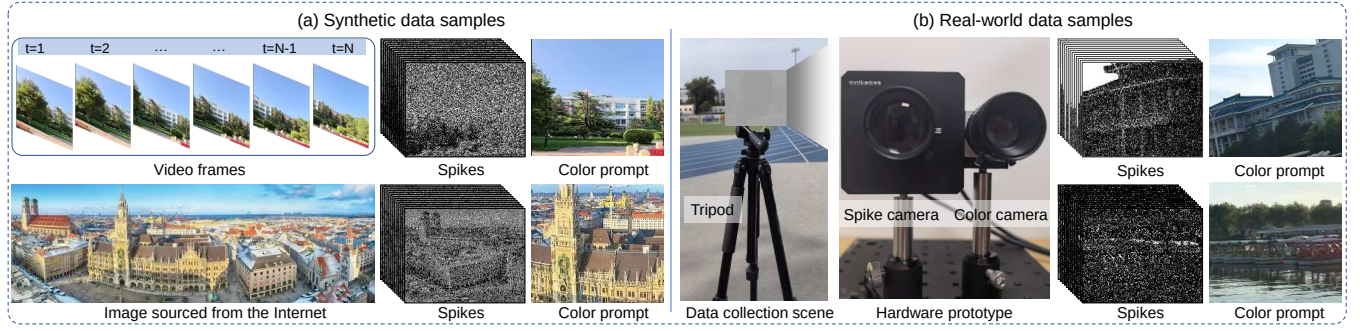


Figure 5: Datasets and hardware prototype. (a) Synthetic data examples. Spike sequences are generated from video frames or panoramic images from Internet, with one RGB image used as the color prompt for fine-tuning. (b) Real-world data examples. The spike camera rotates on a tripod for one second to capture the scene and record the spikes, while a color camera simultaneously captures a single reference color prompt of the same scene.

value is defined as: $a(x, y) = \frac{\langle p - c_1, v \rangle - \min}{\max - \min + \epsilon}$, where $\langle p - c_1, v \rangle$ is the dot product of the vector from p to the geometric center c_1 and the direction vector v , $\min$ and $\max$ are the minimum and maximum values of the projection in the overlapping region, and ϵ is a small constant to avoid division by zero. Therefore, the initial weighted mask S_1 for I_1^w is then given by:

$$S_1 = M'_1 + (1 - a(x, y)) \cdot M_1, \quad (5)$$

where M_1 denotes the overlapping region and M'_1 represents the non-overlapping part, ensuring completeness in non-overlapping areas and smooth gradients in overlaps. We further refine S_1 via joint bilateral filtering:

$$W_1(p) = \frac{1}{k_p} \sum_{q \in \Omega} S_1(q) f(\|p - q\|) g(\|S'_1(p) - S'_1(q)\|), \quad (6)$$

where $f(\cdot)$ and $g(\cdot)$ are spatial and range kernels, and S'_1 is a gaussian-blurred guidance map generated from the initial weighted mask S_1 using gaussian filtering. The resulting weighted mask W_1 enhances texture details while ensuring a smoother transition between overlapping and non-overlapping regions of the image. Finally, we can obtain the panoramic HDR image P :

$$P = \sum_{i=1}^N (W_{i-1} \otimes I_{i-1}^w + W_i \otimes I_i^w). \quad (7)$$

4.3 Generative colorization.

Colorization with only a glance. Considering that most spike cameras are monochromatic, constructing a hybrid camera system similar to Chang et al. 's work [5] is a feasible approach for acquiring color information. However, such systems require strict temporal synchronization and spatial alignment between sensors. To overcome these limitations, we explore whether panoramic HDR colorization can be achieved using only a single-viewpoint (i.e., a glance) color frame as guidance. Inspired by the strong generalization capabilities of large language models, we fine-tune Qwen-Image-Edit-2509 [42] with LoRA [18] on our collected dataset. This approach eliminates the need for strict temporal or spatial alignment between the color camera and the spike camera. As shown in Fig. 4, the color image serves as a visual prompt to provide chromatic cues, enabling our model to perform generative colorization

of the monochromatic panoramic HDR image while preserving fine structural details.

Fine-tuning process. We fine-tune the Qwen-Image-Edit-2509 [42] diffusion transformer using LoRA[18] adapters in the attention projection layers of the DiT [34] module, while keeping the text encoder and VAE [21] frozen. To train, we collect 1,000 images of color panoramas from the internet and convert them into spikes. The reconstructed spike images using TFW serve as edited inputs, with a 256×256 crop from the color image and a short text prompt for visual and textual guidance. Training follows a supervised fine-tuning (SFT) paradigm, with each instance consisting of an edited image, a target image, visual prompt, and optional textual prompt. The model learns to reconstruct the target image conditioned on the edited input under classifier-free guidance. Training is conducted for five epochs with a learning rate of 1×10^{-4} , with the dataset repeated 20 times for stability, taking approximately 72 hours. At inference, both synthetic and real-world spike images are input to the fine-tuned model to generate colorized results. All experiments are performed on a single NVIDIA H20 GPU.

4.4 Dataset

Considering the inherent limitations of spike cameras and the need for scene diversity, we assemble a dataset of synthetic and real-world spikes that supports downstream image reconstruction.

Synthetic data from conventional camera. Capturing diverse real-world scenes is time-consuming and laborious. To enrich scene diversity at low cost, the dataset includes synthetic spikes generated from manually captured video sequences (10–15 seconds each) and panoramic image from the Internet to simulate various scenarios. We construct 9 sequences from video and 1,050 images from diverse sources, with 1,000 of these images used as the training set for fine-tuning. As shown in Fig. 5 (a), the top row shows synthetic samples from video sequences, where consecutive frames are converted into spikes using the strategy in [26], providing a better approximation of real-world spike-camera imaging. Each sequence is paired with a corresponding color image. The bottom row shows samples from single images, with random crops used to simulate spike inputs. A random 256×256 crop is selected as the color prompt, while the full original image serves as the ground-truth panoramic reference.

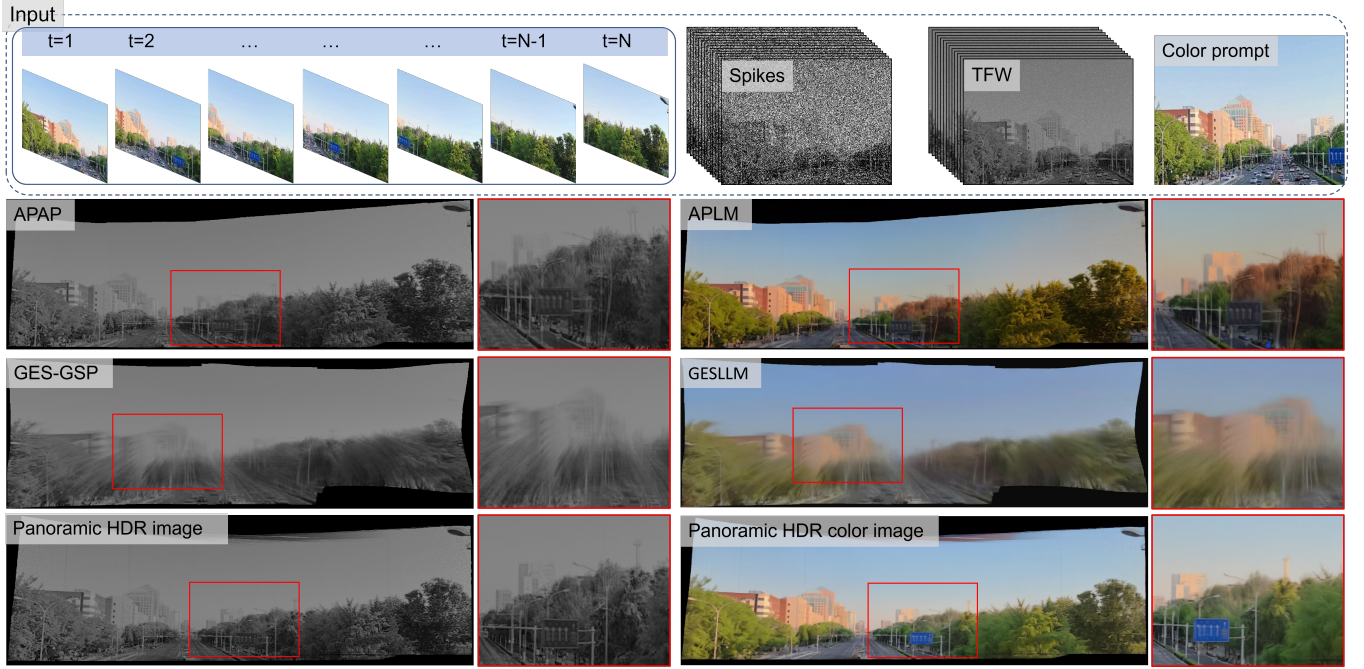


Figure 6: Comparison on synthetic data. We compare our method with the APAP [47] method on monochromatic images and combine APAP [47] with Qwen-Image-Edit-2509 [42] (denoted as APLM) for colorization. Our method effectively reduces motion blur, while the colorization better matches the visual prompt of the color image, resulting in a more natural appearance. *Best viewed by zooming in.*

Table 1: Quantitative results and ablation studies on our synthetic data. $\uparrow$ ($\downarrow$) indicates larger (smaller) values are better. *Ours**: panoramic HDR color image generated by our fine-tuned image editing model.

Method	Comparison on synthetic data						Ablation studies				
	Monochromatic image			Color image			Monochromatic image		Color image		
	APAP [47]	GES-GSP [10]	Ours	APLM	GESLLM	Ours*	w/o enhanced matching	w/o texture-optimized reconstruction	Ours	w/o fine-tuning	Ours*
PSNR $\uparrow$	16.92	16.07	17.52	11.88	11.01	12.45	16.90	17.12	17.52	12.36	12.45
RMSE $\downarrow$	0.15	0.16	0.14	0.26	0.29	0.25	0.15	0.14	0.14	0.37	0.25
HDR-VDP3 $\uparrow$	7.70	7.17	7.83	-	-	-	7.65	7.73	7.83	-	-
DreamSim $\downarrow$	-	-	-	0.22	0.37	0.21	-	-	-	0.22	0.21
$\Delta E00\downarrow$	-	-	-	20.59	22.08	18.19	-	-	-	19.32	18.19

Real-world data from spike camera. To capture diverse scene information for reconstruction in fast-rotating scenes, we exploit the spike camera’s ultra-high sampling rate of 20,000 Hz. As shown in Fig. 5 (b), we collect eight real-world scenes with a spike camera mounted on a tripod, obtaining 50 datasets by rotating the camera 180° for one second to capture raw single-bit spikes. These are decoded into a continuous binary spike stream with a resolution of 250×400 pixels. Meanwhile, a high-resolution color camera is placed side by side with the spike camera to capture a color frame as the prompt.

5 Experiment

5.1 Evaluation on synthetic data

For quantitative evaluation, we use 50 datasets from the Internet, with original images as ground-truth panoramic references. As

shown in Table 1, we compare our method to APAP [47] and GES-GSP [10] on monochromatic images, evaluating PSNR, RMSE and HDR-VDP3 [28]. To ensure fairness, ground-truth images are converted into spikes using the strategy in [26] and reconstructed via TFW. Our method performs better. For color evaluation, we further compare APAP [47] combined with Qwen-Image-Edit-2509 [42], denoted as APLM, and GES-GSP [10] combined with Qwen-Image-Edit-2509 [42], denoted as GESLLM. Performance is evaluated using PSNR, RMSE, DreamSim [15] (for perceptual similarity), and $\Delta E00$ [37] (for color fidelity). Using the prompt “Colors in Image 1 are adjusted to match the style of the color image in Image 2”, our fine-tuned model produces superior performance on the color images.

For qualitative analysis, we compare our method with APAP [47] on monochromatic images and APLM on color images. As shown in Fig. 6, APAP produces noticeable blur (highlighted by red boxes),

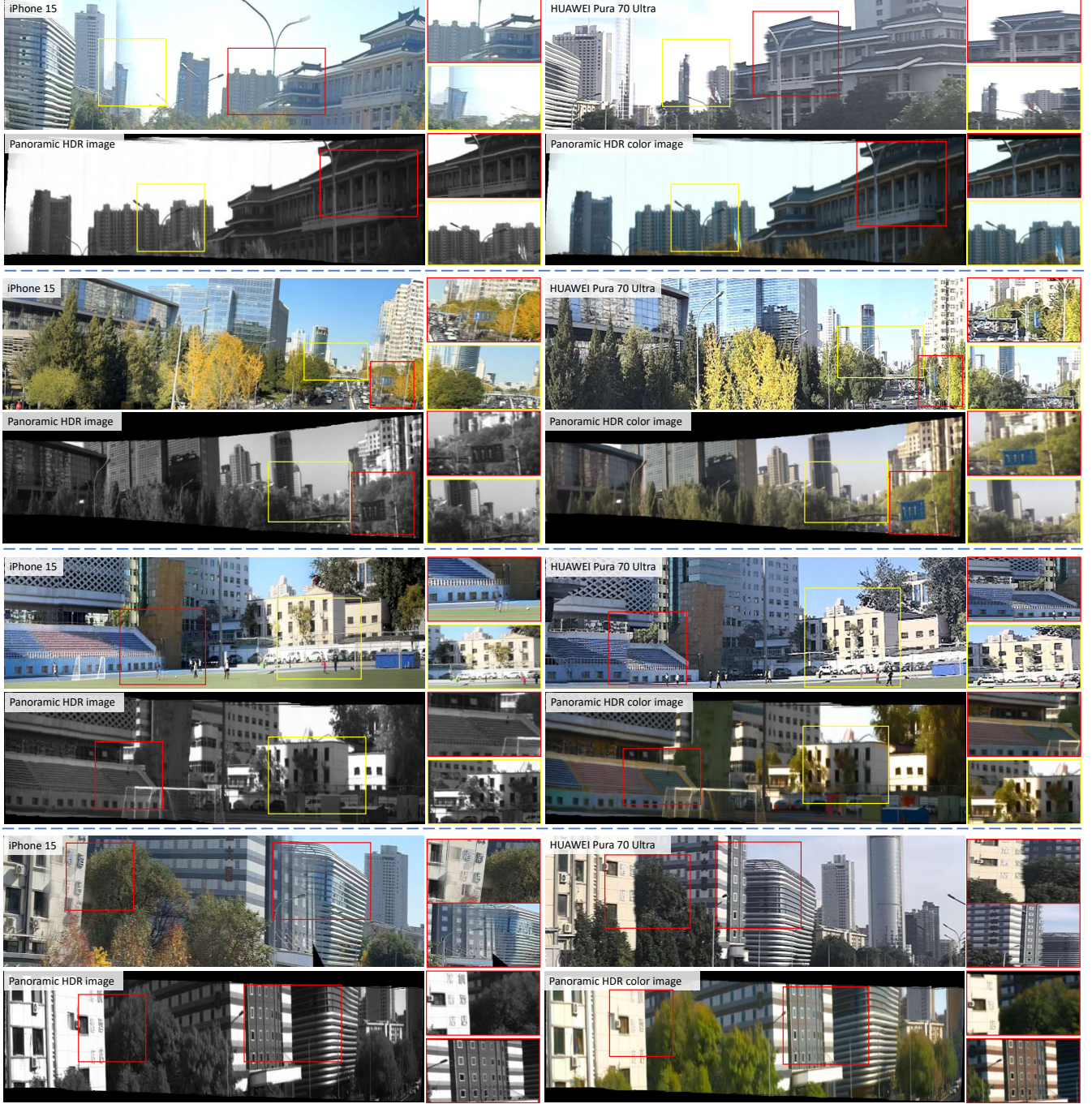


Figure 7: Comparison of real-world data. We compare panoramic results captured by the iPhone 15 and HUAWEI Pura 70 Ultra using their built-in panorama modes for the same scene. Results obtained by both smartphones exhibit noticeable motion artifacts (red boxes) in fast-rotating scenes and overexposure (yellow boxes) in high dynamic range regions. *Best viewed by zooming in.*

while our method achieves clearer textures. APLM suffers from motion blur, and the generated colors do not match the visual prompt of the color image as well, resulting in a less natural appearance compared to our method.

5.2 Evaluation on real-world data

We acknowledge that comparing our method with those in the table may not be entirely fair. Therefore, in fast-rotating scenes, we further compare it with the panoramic modes of conventional



Figure 8: Ablation studies on our DSFT-enhanced vertical constraint matching. Without the proposed DSFT-enhanced feature extraction and the vertical distance constraint, the reconstruction fails due to inaccurate feature correspondences. In contrast, ours enables accurate correspondences and successful panorama reconstruction.

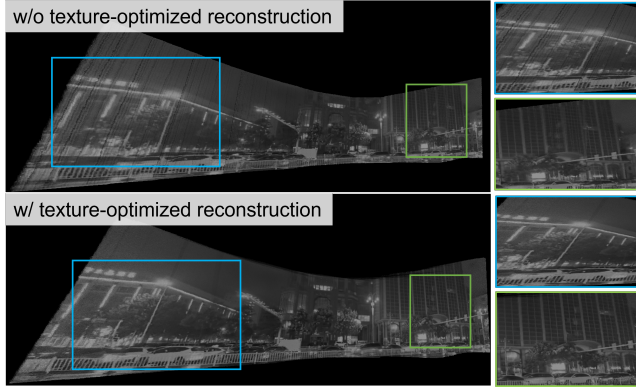


Figure 9: Ablation studies on our texture-optimized reconstruction. Without the weighted mask, noticeable seam boundaries (blue boxes) and visible artifacts (green boxes) appear. In contrast, our method effectively suppresses these issues and preserves fine textures. Best viewed by zooming in.

cameras, including the iPhone 15 and HUAWEI Pura 70 Ultra, in real-world data. The reconstructed images from the spike camera generally exhibit low contrast and dark intensity distribution. To facilitate color filling and detail restoration in the fine-tuned model, gamma correction with $\gamma = 1.5$ is applied to enhance brightness and contrast. As shown in Fig. 7, in the first scene, the iPhone 15 loses part of the building (red box) and introduces ghosting from overexposure (yellow box). The HUAWEI Pura 70 Ultra shows roof misalignment (red box) and overexposure that destroys structural details (yellow box). In the second scene, the signboard suffers motion blur (red box), and distant buildings are overexposed to white (yellow box). In contrast, our method reconstructs a panoramic HDR color image from spike data under fast rotation and high dynamic range conditions.

5.3 Ablation study

We conduct ablation studies to assess the effectiveness of each module in our framework. First, we evaluate the spike-based reconstruction on 50 synthetic monochromatic datasets, with ground-truth images converted for consistency. For “w/o DSFT-enhanced vertical constraint matching”, as shown in Table 1, due to inherent spike noise affecting feature matching, our method enhances feature points using DSFT and a vertical distance constraint to reduce matching errors, improving image reconstruction quality.

We further evaluate visual quality on real-world data, as shown in Fig. 8. Our method successfully avoids reconstruction failure, enabling visually consistent panoramic reconstruction. For “w/o texture-optimized reconstruction”, as shown in Fig. 9, without the masks, accumulated artifacts (green boxes) and visible seams (blue boxes) appear, while our approach preserves finer textures and improves visual quality. Finally, for “w/o fine-tuning”, we apply the web-based image editing model [42] with the prompt “Colors in Image 1 are adjusted to match the style of the color image in Image 2.” As shown in Table 1, fine-tuning produces more perceptually consistent color distributions, enhancing visual quality.

6 Conclusion

In this paper, we present a new framework for reconstructing panoramic HDR color image from single-second spikes. We first propose a spike-based panoramic reconstruction method to generate panoramic HDR images. Taking into account the noise inherent in spike data and leveraging the differential spike firing time (DSFT) to enhance the sharpness of fine details, we design a DSFT-enhanced vertical constraint matching and texture optimization module. This module significantly improves reconstruction quality by suppressing noise while reducing stitching artifacts in overlapping regions. Considering that spike cameras inherently lack color information due to their binary sensing mechanism, we employ a side-by-side color camera to capture a color frame as a visual prompt, fine-tuning an image editing model to generate a panoramic HDR color image in fast-rotating scenes.

Limitations and future work. We acknowledge that large generative models may introduce hallucinations when inputs deviate from the training distribution. However, spike-based texture reconstruction and learned color priors provide strong constraints, helping preserve geometric and textural consistency while enabling plausible color transfer. In practice, hallucinations are mostly limited to reasonable color variations rather than structural distortions. In addition, the relatively low resolution of the spike camera constrains reconstruction quality. To address these challenges, future work will explore higher-resolution devices, such as the Spike M1K40-H2-Gen3, to fundamentally improve image sharpness and visual quality. Additionally, we aim to integrate neural rendering techniques to further enhance dynamic range performance in complex real-world environments.

Acknowledgments

This work was supported by National Key R&D Program of China (2024YFB3909900), Beijing High Innovation Plan (Contract No. 202604841443), and Beijing Natural Science Foundation (Grant No. L233024).

References

- [1] Yunpeng Bai, Chao Dong, Zenghao Chai, Andong Wang, Zhengzhuo Xu, and Chun Yuan. 2022. Semantic-sparse colorization network for deep exemplar-based colorization. *arXiv:2112.01335 [cs.CV]* <https://arxiv.org/abs/2112.01335>
- [2] Matthew Brown and David G Lowe. 2007. Automatic panoramic image stitching using invariant features. *International Journal of Computer Vision* 74 (2007), 59–73.
- [3] Yakun Chang, Zhaojun Huang, Siqi Yang, Yeliduosi Xiaokaiti, Shikui Wei, Yao Zhao, Tiejun Huang, and Boxin Shi. 2026. HFR and HDR video from multi-attenuated spikes using a rapidly rotating spokeND filter. In *IEEE Conf. Comput. Vis. Pattern Recog.* 19716–19725.
- [4] Yakun Chang, Yeliduosi Xiaokaiti, Yujia Liu, Bin Fan, Zhaojun Huang, Tiejun Huang, and Boxin Shi. 2024. Towards HDR and HFR video from rolling-mixed-bit spikings. In *IEEE Conf. Comput. Vis. Pattern Recog.* 25117–25127.
- [5] Yakun Chang, Chu Zhou, Yuchen Hong, Liwen Hu, Chao Xu, Tiejun Huang, and Boxin Shi. 2023. 1000 fps hdr video with a spike-rgb hybrid camera. In *IEEE Conf. Comput. Vis. Pattern Recog.* 22180–22190.
- [6] Zheng Chang, Shuchen Weng, Yu Li, Si Li, and Boxin Shi. 2022. L-CoDer: Language-based colorization with color-object decoupling transformer. In *Eur. Conf. Comput. Vis.*
- [7] Zheng Chang, Shuchen Weng, Peixuan Zhang, Yu Li, Si Li, and Boxin Shi. 2023. L-CoIns: Language-based colorization with instance awareness. In *IEEE Conf. Comput. Vis. Pattern Recog.*
- [8] Padraig Cunningham and Sarah Jane Delany. 2021. K-nearest neighbour classifiers-a tutorial. *ACM computing surveys (CSUR)* 54, 6 (2021), 1–25.
- [9] Yanchen Dong, Ruiqin Xiong, Jing Zhao, Jian Zhang, Xiaopeng Fan, Shuyuan Zhu, and Tiejun Huang. 2024. Joint demosaicing and denoising for spike camera. In *AAAI*, Vol. 38. 1582–1590.
- [10] Peng Du, Jifeng Ning, Jiguang Cui, Shaoli Huang, Xinchao Wang, and Jiaxin Wang. 2022. Geometric structure preserving warp for natural image stitching. In *IEEE Conf. Comput. Vis. Pattern Recog.* 3688–3696.
- [11] Ashley Eden, Matthew Uyttendaele, and Richard Szeliski. 2006. Seamless image stitching of scenes with large motions and exposure differences. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, Vol. 2. IEEE, 2498–2505.
- [12] Johan Edstedt, Qiyu Sun, Georg Bökman, Márten Wadenbäck, and Michael Felsberg. 2024. Roma: Robust dense feature matching. In *IEEE Conf. Comput. Vis. Pattern Recog.* 19790–19800.
- [13] Zeev Farberman, Gil Hoffer, Yaron Lipman, Daniel Cohen-Or, and Dani Lischinski. 2009. Coordinates for instant image cloning. *ACM Transactions on graphics (TOG)* 28, 3 (2009), 1–9.
- [14] Martin A Fischler and Robert C Bolles. 1981. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM* 24, 6 (1981), 381–395.
- [15] Stephanie Fu, Netanel Tamir, Shobhita Sundaram, Lucy Chai, Richard Zhang, Tali Dekel, and Phillip Isola. 2023. Dreamsim: Learning new dimensions of human visual similarity using synthetic data. *arXiv preprint arXiv:2306.09344* (2023).
- [16] Jin Han, Chu Zhou, Peiqi Duan, Yehui Tang, Chang Xu, Chao Xu, Tiejun Huang, and Boxin Shi. 2020. Neuromorphic camera guided high dynamic range imaging. In *IEEE Conf. Comput. Vis. Pattern Recog.*
- [17] Mingming He, Dongdong Chen, Jing Liao, Pedro V Sander, and Lu Yuan. 2018. Deep exemplar-based colorization. *ACM Transactions on Graphics (TOG)* 37, 4 (2018), 47.
- [18] Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, Weizhu Chen, et al. 2022. Lora: Low-rank adaptation of large language models. *ICLR* 1, 2 (2022), 3.
- [19] Tiejun Huang, Yajing Zheng, Zhaofei Yu, Rui Chen, Yuan Li, Ruiqin Xiong, Lei Ma, Junwei Zhao, Siwei Dong, Lin Zhu, et al. 2022. 1000× faster camera and machine vision with ordinary devices. *Engineering* (2022).
- [20] Qi Jia, Xiaomei Feng, Yu Liu, Xin Fan, and Longjin Jan Latecki. 2023. Learning pixel-wise alignment for unsupervised image stitching. In *Proceedings of the 31st ACM International Conference on Multimedia*. 1392–1400.
- [21] Diederik P Kingma and Max Welling. 2013. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114* (2013).
- [22] Anat Levin, Assaf Zomet, Shmuel Peleg, and Yair Weiss. 2004. Seamless image stitching in the gradient domain. In *Eur. Conf. Comput. Vis.* Springer, 377–389.
- [23] Hao Liang, Zhipeng Dong, Kaixin Chen, Jiyuan Guo, Yufeng Yue, Yi Yang, and Mengyin Fu. 2025. ChatStitch: Visualizing through structures via surround-view unsupervised deep image stitching with collaborative LLM-Agents. *arXiv preprint arXiv:2503.14948* (2025).
- [24] Tianli Liao and Nan Li. 2019. Single-perspective warps in natural image stitching. *IEEE transactions on image processing* 29 (2019), 724–735.
- [25] Tony Lindeberg. 2012. Scale invariant feature transform. (2012).
- [26] Lei Ma, Zhaofei Yu, Boxin Shi, Tiejun Huang, et al. 2023. Spike stream denoising via spike camera simulation. *arXiv preprint arXiv:2304.03129* (2023).
- [27] Varun Manjunatha, Mohit Iyyer, Jordan Boyd-Graber, and Larry Davis. 2018. Learning to Color from Language. In *North american chapter of the association for computational linguistics*.
- [28] Rafal K. Mantiuk, Dounia Hammou, and Param Hanji. 2023. HDR-VDP-3: A multi-metric for predicting image differences, quality and contrast distortions in high dynamic range and regular content. *arXiv:2304.13625 [eess.IV]* <https://arxiv.org/abs/2304.13625>
- [29] Mingyuan Meng and Shaojun Liu. 2020. High-quality panorama stitching based on asymmetric bidirectional optical flow. In *2020 5th International Conference on Computational Intelligence and Applications (ICCIA)*. IEEE, 118–122.
- [30] Lang Nie, Chunyu Lin, Kang Liao, Meiqin Liu, and Yao Zhao. 2020. A view-free image stitching network based on global homography. *Journal of Visual Communication and Image Representation* 73 (2020), 102950.
- [31] Lang Nie, Chunyu Lin, Kang Liao, Shuaicheng Liu, and Yao Zhao. 2021. Unsupervised deep image stitching: Reconstructing stitched features to images. *IEEE Transactions on Image Processing* 30 (2021), 6184–6197.
- [32] Lang Nie, Chunyu Lin, Kang Liao, Yun Zhang, Shuaicheng Liu, and Yao Zhao. 2025. StabStitch+: Unsupervised online video stitching with spatiotemporal bidirectional warps. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2025).
- [33] Lang Nie, Yuan Mei, Kang Liao, Yunqiu Xu, Chunyu Lin, and Bin Xiao. 2025. Robust image stitching with optimal plane. *arXiv preprint arXiv:2508.05903* (2025).
- [34] William Peebles and Saining Xie. 2023. Scalable diffusion models with transformers. In *Int. Conf. Comput. Vis.* 4195–4205.
- [35] Zhong Qu, Jun Li, Kang-Hua Bao, and Zhi-Chao Si. 2020. An unordered image stitching method based on binary tree and estimated overlapping area. *IEEE Transactions on Image Processing* 29 (2020), 6734–6744.
- [36] Zi Siang See, Lizbeth Goodman, Craig Hight, Mohd Shahrizal Sunar, Arindam Dey, Yen Kaow Ng, and Mark Billinghurst. 2019. Creating high fidelity 360° degree virtual reality with high dynamic range spherical panorama images. (2019).
- [37] Gaurav Sharma, Wencheng Wu, and Edul N Dalal. 2005. The CIEDE2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations. *Color Research & Application: Endorsed by Inter-Society Color Council, The Colour Group (Great Britain), Canadian Society for Color, Color Science Association of Japan, Dutch Society for the Study of Color, The Swedish Colour Centre Foundation, Colour Society of Australia, Centre Français de la Couleur* 30, 1 (2005), 21–30.
- [38] Dae-Young Song, Geonsoo Lee, HeeKyung Lee, Gi-Mun Um, and Donghyeon Cho. 2022. Weakly-supervised stitching network for real-world panoramic image generation. In *Eur. Conf. Comput. Vis.* Springer, 54–71.
- [39] Mathieu Tuli, Kaveh Kamali, and David B Lindell. 2025. Generative Panoramic Image Stitching. *arXiv preprint arXiv:2507.07133* (2025).
- [40] Ziyu Wan, Bo Zhang, Dongdong Chen, Pan Zhang, Dong Chen, Jing Liao, and Fang Wen. 2020. Bringing old photos back to life. In *IEEE Conf. Comput. Vis. Pattern Recog.* 2747–2757.
- [41] Shuchen Weng, Hao Wu, Zheng Chang, Jiajun Tang, Si Li, and Boxin Shi. 2022. L-CoDe: Language-based colorization using color-object decoupled conditions. In *AAAI*.
- [42] Chenfei Wu, Jiahao Li, Jingren Zhou, Junyang Lin, Kaiyuan Gao, Kun Yan, Shengming Yin, Shuai Bai, Xiao Xu, Yilei Chen, et al. 2025. Qwen-image technical report. *arXiv preprint arXiv:2508.02324* (2025).
- [43] Yi Xiao, Peiyao Zhou, Yan Zheng, and Chi-Sing Leung. 2019. Interactive deep colorization using simultaneous global and local inputs. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 1887–1891.
- [44] Yingen Xiong and Kari Pulli. 2009. Mask-based image blending and its applications on mobile devices. In *MIPPR 2009: Remote Sensing and GIS Data Processing and Other Applications*, Vol. 7498. SPIE, 1066–1073.
- [45] Siqi Yang, Zhaojun Huang, Yakun Chang, Bin Fan, Zhaofei Yu, and Boxin Shi. 2024. Real-data-driven 2000 FPS color video from mosaicked chromatic spikes. In *Eur. Conf. Comput. Vis.* 1111–2222.
- [46] Jooyeol Yun, Sanghyeon Lee, Minho Park, and Jaegul Choo. 2023. iColoriT: Towards propagating local hints to the right region in interactive colorization by leveraging vision transformer. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*. 1787–1796.
- [47] Julio Zaragoza, Tat-Jun Chin, Michael S Brown, and David Suter. 2013. As-projective-as-possible image stitching with moving DLT. In *IEEE Conf. Comput. Vis. Pattern Recog.* 2339–2346.
- [48] Guofeng Zhang, Yi He, Weifeng Chen, Jiaya Jia, and Hujun Bao. 2016. Multi-viewpoint panorama construction with wide-baseline images. *IEEE Transactions on Image Processing* 25, 7 (2016), 3099–3111.

- [49] Jing Zhao, Ruiqin Xiong, Hangfan Liu, Jian Zhang, and Tiejun Huang. 2021. Spk2ImgNet: Learning to reconstruct dynamic scene from continuous spike stream. In *IEEE Conf. Comput. Vis. Pattern Recog.* 11996–12005.
- [50] Jing Zhao, Ruiqin Xiong, Jiyu Xie, Boxin Shi, Zhaoifei Yu, Wen Gao, and Tiejun Huang. 2021. Reconstructing clear image for high-speed motion scene with a retina-inspired spike camera. 8 (2021), 12–27.
- [51] Jing Zhao, Ruiqin Xiong, Jian Zhang, Rui Zhao, Hangfan Liu, and Tiejun Huang. 2023. Learning to super-resolve dynamic scenes for neuromorphic spike camera. In *AAAI*, Vol. 37. 3579–3587.
- [52] Junwei Zhao, Jianming Ye, Shiliang Shiliang, Zhaoifei Yu, and Tiejun Huang. 2023. Recognizing high-speed moving objects with spike camera. 7657–7665.
- [53] Rui Zhao, Ruiqin Xiong, Jing Zhao, Zhaoifei Yu, Xiaopeng Fan, and Tiejun Huang. 2022. Learning optical flow from continuous spike streams. *Adv. Neural Inform. Process. Syst.* 35 (2022), 7905–7920.
- [54] Lin Zhu, Siwei Dong, Tiejun Huang, and Yonghong Tian. 2019. A retina-inspired sampling method for visual texture reconstruction. In *Int. Conf. Multimedia and Expo*. 1432–1437.
- [55] Lin Zhu, Siwei Dong, Jianing Li, Tiejun Huang, and Yonghong Tian. 2020. Retina-like visual image reconstruction via spiking neural model. In *IEEE Conf. Comput. Vis. Pattern Recog.* 1438–1446.

Fast-Rotating Panoramic HDR Reconstruction from Single-Second Spikes

Yuzhu Tao*

Beijing Jiaotong University
Beijing, China
taoyuzhu@bjtu.edu.cn

Shikui Wei*

Beijing Jiaotong University
Beijing, China
shkwei@bjtu.edu.cn

Lang Nie

Chongqing University of Posts and
Telecommunications
Chongqing, China
nielang@cqupt.edu.cn

Qingwen Li

China Tower Corporation Limited
Beijing, China
liqw3@chinatowercom.cn

Bo Li

China Tower Corporation Limited
Beijing, China
libo88872@chinatowercom.cn

Boxin Shi[†]

Peking University
Beijing, China
shiboxin@pku.edu.cn

Yao Zhao*

Beijing Jiaotong University
Beijing, China
yzhao@bjtu.edu.cn

Yakun Chang^{‡*}

Beijing Jiaotong University
Beijing, China
ykchang@bjtu.edu.cn

In this supplementary material, we provide details of our dataset in Sec. 1, details of reconstruction in Sec. 2 and more results in Sec. 3.

1 More Details of Dataset

Synthetic data. We convert consecutive frames captured from video sequences and randomly cropped images from Internet panoramic images into spike streams using the noise modeling strategy in [4], producing sequences with intrinsic spatiotemporal noise that approximate real spike-camera outputs. We then reconstruct images from these synthetic spikes. Since the time window used in texture-from-window (TFW) [7] reconstruction can introduce noise or blur if set too small or too large, our experiments show that a window size of $M = 20$ provides a good balance on synthetic data. In addition, incorporating differential spike firing time (DSFT) [6] further preserves fine texture details during reconstruction.

Real-world data. Spike cameras [1–3] have shown significant promise for high-speed applications. In this study, we use a spike camera mounted on a tripod to capture fast-rotating scenes, rotating 180° over 1 second, generating 20,000 frames across 50 data files, each containing 400 frames. Each frame is represented by a single-bit spike event, capturing discrete moments of the scene with high temporal resolution. To reconstruct the images, we set the texture frame window (TFW) to $M = 50$, meaning every 50 accumulated spikes generate a single image, resulting in 400 images from the 20,000 frames. These images are then stitched together to form the full panoramic scene. The advantage of using a spike-based camera lies in its ability to capture high dynamic range (HDR) information and handle rapid rotational motion, challenges for conventional cameras due to their slower frame rates and limited resolution.

*Institute of Information Science, Beijing Jiaotong University; Visual Intelligence +X International Cooperation Joint Laboratory of the MoE.

[†]State Key Laboratory of Multimedia Information Processing, School of Computer Science, Peking University; National Engineering Research Center of Visual Technology, School of Computer Science, Peking University; and PKU-AI² Robotics Joint Lab of Embodied AI.

[‡]Corresponding author.

This approach enables efficient image reconstruction in HDR, fast-rotating environments, producing high-quality panoramic imagery from sparse data.

2 More Details of Reconstruction

Considering the impact of noise inherent in the spike camera itself, we incorporated a DSFT-enhanced vertical constraint matching framework in the feature matching process to improve the accuracy of the matched points. To better visualize the matching process, as shown in Fig. 12, we observe in (a) that some incorrect matches are indicated by the red arrows. In (b), the number of incorrect matches is significantly reduced, and we can also see that additional matches have been successfully identified in the yellow box, thanks to the DSFT-based image matching enhancement. This improvement results in more precise matching points, making image alignment significantly easier. Since spike cameras generate dense frames, resulting in the reconstruction of a large number of images, directly stitching them in sequence inevitably leads to the accumulation of geometric errors. To mitigate this, we organize the image sequence into a balanced binary partition tree. The stitching process then proceeds in a bottom-up fashion, with the left and right branches growing and converging near the midpoint. This strategy effectively suppresses drift, ensuring the reconstruction of a geometrically consistent and visually natural panoramic image.

During the synthesis phase, we employ an adaptive transformation—comprising a global transformer and a mesh-based local transformation (MBLT)[5] to align the input images. The aligned images are then composited into a common coordinate system using weighted masks. For the MBLT component, we adopt the off-the-shelf implementation provided in [5]. To verify its effectiveness in our specific setting, we conduct additional experiments on real-world scenes. As shown in Fig. 14, removing MBLT introduces noticeable black stripes, whereas our full pipeline yields substantially better alignment and reconstruction quality. Fig. 13

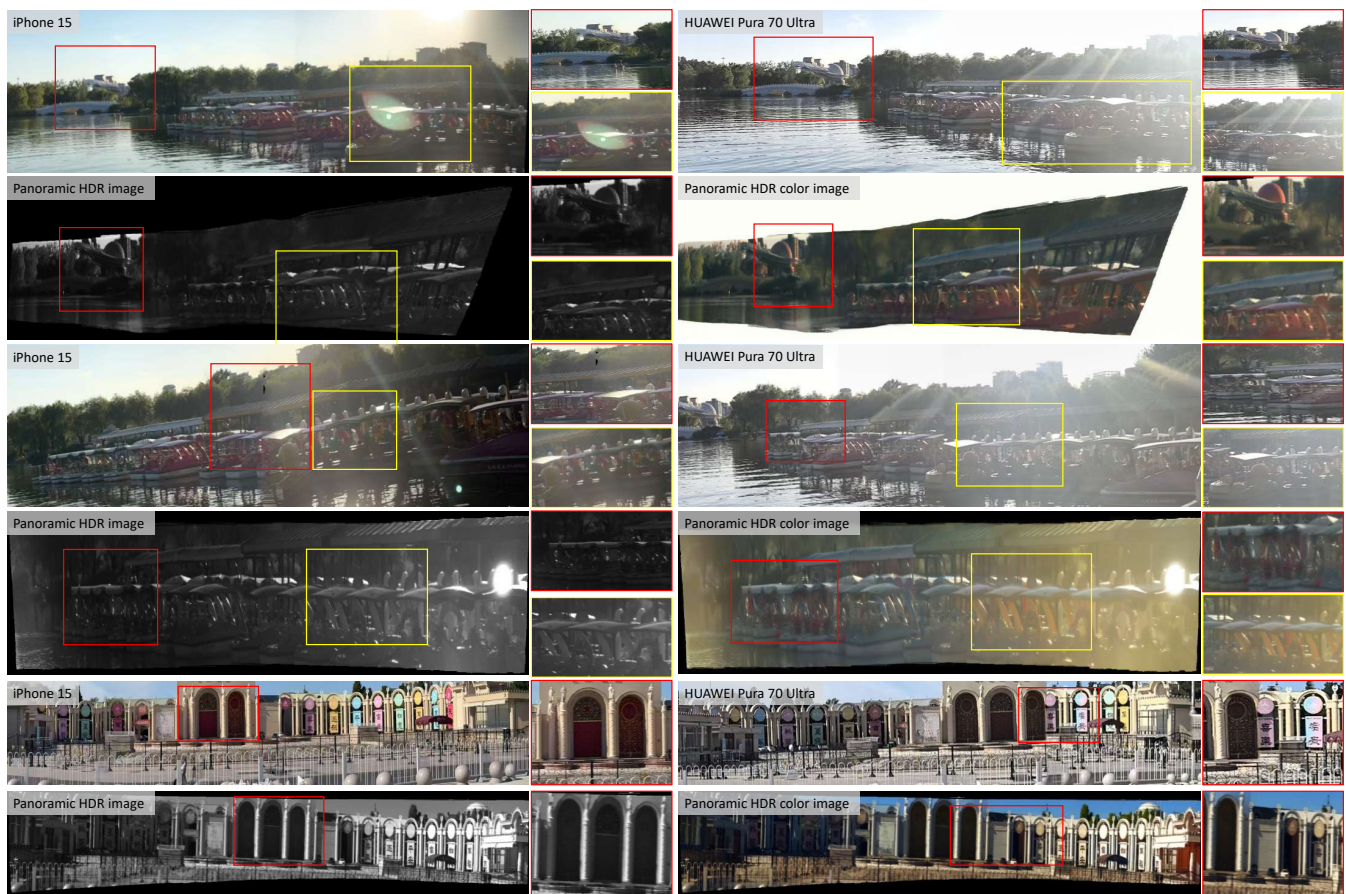


Figure 10: Comparative results across diverse scenarios.



Figure 11: More results from our real-world data using our method for reconstruction with the hybrid camera system.

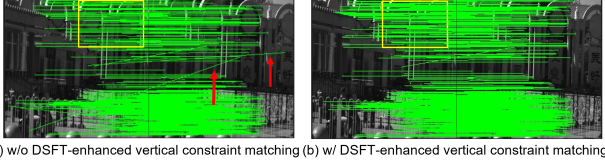


Figure 12: A visualization of our DSFT-enhanced vertical constraint matching on an image pair. Red arrows indicate matches with significant vertical disparity, while yellow boxes highlight additional correct correspondences introduced by the DSFT enhancement.



Figure 13: The final fusion step. Row 1 shows the weighted masks for the two input images. Row 2 presents the images aligned into a common coordinate frame. Row 3 displays the resulting panoramic HDR image and panoramic HDR color image.

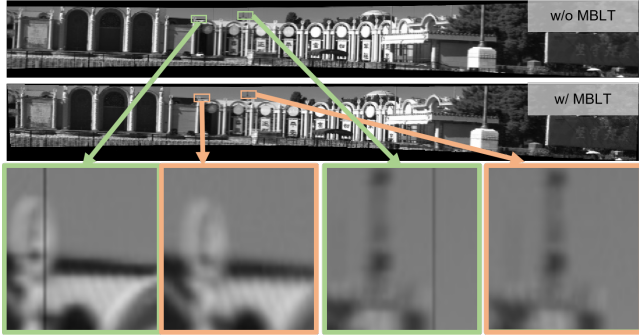


Figure 14: Effect of the mesh-based local transformation (MBLT). Removing MBLT leads to noticeable black stripes (left), whereas our full pipeline with MBLT achieves substantially better alignment and reconstruction (right).

illustrates the estimated weighted masks computed during the final fusion step, along with the corresponding source images. In overlapping regions, we apply weight-based fusion to minimize artifacts, ultimately producing a seamless panoramic output.

3 More Details of Experiments

To comprehensively evaluate image quality in real-world scenarios, we illustrate more results across diverse scenarios, comparing the performance of iPhone 15 and Huawei Pura 70 Ultra under identical conditions. As shown in Fig. 10, our proposed method is capable of reconstructing high-quality panoramic HDR images in fast-rotating scenes. We utilize a side-by-side color camera system to capture a reference color frame, which serves as a color prompt for the subsequent image reconstruction. The color frame from the

conventional camera is used to guide the fine-tuning of an image-editing model, specifically optimized for HDR color reproduction. Meanwhile, we also present more results obtained from real-world data. As shown in Fig. 11, we showcase panoramic HDR images and their corresponding panoramic HDR color images, captured from additional scenes using our method. The resulting panoramic HDR color images closely match the color representation of real-world scenes, yielding visually pleasing results that are consistent with human visual perception.

References

- [1] Yakun Chang, Chu Zhou, Yuchen Hong, Liwen Hu, Chao Xu, Tiejun Huang, and Boxin Shi. 2023. 1000 fps hdr video with a spike-rgb hybrid camera. In *IEEE Conf. Comput. Vis. Pattern Recog.* 22180–22190.
- [2] Jin Han, Chu Zhou, Peiqi Duan, Yehui Tang, Chang Xu, Chao Xu, Tiejun Huang, and Boxin Shi. 2020. Neuromorphic camera guided high dynamic range imaging. In *IEEE Conf. Comput. Vis. Pattern Recog.*
- [3] Tiejun Huang, Yajing Zheng, Zhaofei Yu, Rui Chen, Yuan Li, Ruiqin Xiong, Lei Ma, Junwei Zhao, Siwei Dong, Lin Zhu, et al. 2022. 1000× faster camera and machine vision with ordinary devices. *Engineering* (2022).
- [4] Lei Ma, Zhaofei Yu, Boxin Shi, Tiejun Huang, et al. 2023. Spike stream denoising via spike camera simulation. *arXiv preprint arXiv:2304.03129* (2023).
- [5] Julio Zaragoza, Tat-Jun Chin, Michael S Brown, and David Suter. 2013. As-projective-as-possible image stitching with moving DLT. In *IEEE Conf. Comput. Vis. Pattern Recog.* 2339–2346.
- [6] Rui Zhao, Ruiqin Xiong, Jing Zhao, Zhaofei Yu, Xiaopeng Fan, and Tiejun Huang. 2022. Learning optical flow from continuous spike streams. *Adv. Neural Inform. Process. Syst.* 35 (2022), 7905–7920.
- [7] Lin Zhu, Siwei Dong, Jianing Li, Tiejun Huang, and Yonghong Tian. 2020. Retina-like visual image reconstruction via spiking neural model. In *IEEE Conf. Comput. Vis. Pattern Recog.* 1438–1446.