E2PRIV: Privacy-Preserving Event-to-Video Reconstruction with Face Anonymization

1st Mira Adra
GTD International
Toulouse, France
mira.adra@gtd.eu

2nd Jean-Luc Dugelay
Digital Security Department
Eurecom
Biot, France
jld@eurecom.fr

***Abstract*—Event cameras are increasingly adopted in human-centered applications due to their ability to capture high-speed motions without blur and fine-grained details such as facial micro-expressions. While their unconventional pixel-based data structure is often considered privacy-preserving, event-to-video reconstruction algorithms challenge this assumption by revealing identifiable faces from event streams, commonly known as a visualization attack. Since these sensors offer high temporal and dynamic resolution, reconstructed faces often retain fine details, increasing the risk of sensitive information leakage. To address this challenge, anonymization techniques have been explored, but they are typically applied post hoc as independent processing steps. Accordingly, we propose E2PRIV, the first event-based reconstruction model that integrates anonymization directly within the reconstruction process. By incorporating dynamic Gaussian blurring and noise injection during training, our model effectively obscures identifiable facial features while balancing the trade-off between privacy and usability. We evaluate our anonymization approach through face detection and verification, demonstrating a significant reduction in both detected faces and identity matching rates. To further validate its effectiveness, we study the impact of anonymization on action recognition and prove that anonymized frames maintain performance without degradation, ensuring that they remain functional for human-centered downstream applications. Code available at: https://github.com/MyraAdra/e2priv**

***Index Terms*—Event Cameras, privacy-preserving, event anonymization, face recognition, security**

I. Introduction

Unlike traditional cameras that generate RGB frames capable of revealing sensitive information, event cameras produce data as sparse and asynchronous updates based on brightness variations at pixel level. This unconventional format has led to their integration into face-related applications, such as eye-blink authentication, where they capture fine-grained facial movements while minimizing concerns about identity exposure. Accordingly, event cameras, also called dynamic vision sensors, have gained widespread use in the domain of video surveillance and security due to the assumption that they are privacy-preserving by nature considering their data sparsity and lack of visual detail.

Several studies have even proposed so-called "privacy-preserving models" solely because they use event data [1], [2]. Interestingly, this belief has also inspired algorithms that simulate event-based data for privacy purposes. For instance, Kumawat et al. [3] introduced the BDQ encoder, a digital approximation of event cameras that transforms RGB inputs into event-based representations, enabling privacy-aware processing in the frame-based domain.

However, recent studies have demonstrated that event streams are vulnerable to privacy breaches not only during transmission but also at various stages of processing. Zhang et al. categorize these attacks as either recognition or visualization attacks [4]. For example, advanced reconstruction algorithms, such as E2VID [5], [6], can invert event data to generate high-fidelity images, including identifiable faces, effectively turning event cameras into tools for inversion attacks. Moreover, in contrast to traditional cameras, where low light or motion blur can mask identities, the high dynamic range and fine temporal granularity of event cameras preserve sharp facial details, significantly amplifying privacy risks upon reconstruction.

This challenge has driven the development of anonymization techniques to address privacy risks while ensuring the suitability of event cameras for high-security and human-centered applications. However, many approaches treat reconstruction algorithms solely as an attack vector, overlooking their potential to improve event data processing and performance. In reality, reconstructed frames often outperform raw event data by leveraging the powerful feature extraction capabilities of Convolutional Neural Networks (CNNs) [5]. While anonymization strategies are well-established for traditional RGB systems, there remains a significant gap in addressing face anonymization for event streams. To date, existing solutions fail to ensure that reconstructed frames are both anonymized and functional, often compromising privacy at the expense of usability or vice versa [7]. In traditional RGB-based systems, anonymization is applied after frames are already generated—whether through blurring, masking, or generative models—leaving the original, non-anonymized data accessible. Similarly, event-based approaches often introduce anonymization modules that operate on raw event streams before further processing. However, these methods distort the event data, making it unsuitable for reconstruction-based applications. Since event reconstruction remains one of the most effective ways to leverage event data for CNN-based tasks, our approach integrates anonymization at this stage, ensuring that the first visible representation of the data is already anonymized while maintaining usability.

Unlike conventional post-processing methods, this approach does not alter the original event stream but prevents reversible transformations on the reconstructed frames.

To overcome this limitation, in this paper, we propose a privacy-by-design approach that integrates anonymization directly into the reconstruction process in an irreversible manner. Our method transitions directly from raw event data to anonymized frames, ensuring that identifiable facial features are never exposed. This ensures that event cameras can be securely used in sensitive applications, such as video surveillance and driver monitoring while conforming to security guidelines and preventing the leakage of sensitive facial information. Our contributions can be summarized as follows:

- To the best of our knowledge, we propose E2PRIV as the first event-based reconstruction model that distorts faces in event streams by re-training the E2VID architecture with dynamic blurring and noise injection on a face dataset.
- We evaluate the model on benchmark datasets using face detection and verification as criteria, demonstrating a significant reduction in both detection and re-identification rates.
- We demonstrate that anonymized event reconstruction does not compromise performance in human-centered downstream applications, such as action recognition, and can be effectively used as a privacy-preserving substitute.

II. Related Works

A. Event Data Reconstruction

Event cameras produce asynchronous and sparse pixel-level data, requiring specialized event representations to process the data effectively with existing networks and architectures. One approach is to create frame-based representations that allow event data to be processed with traditional CNNs. Among these, reconstructing intensity frames using models like E2VID has proven to be the better-performing method, as it fully leverages the power of CNNs for event-based downstream applications.

One of the first works to prove that event-to-image reconstruction is possible is the one proposed by Cook et al. [8] and Kim et al. [9]. These, among other early algorithms, depended largely on leveraging camera motion and optical flow to estimate the intensity and were limited to reconstructing a single frame rather than a video. Recent advancements, however, have shifted towards direct integration of events, allowing for higher-frame rate reconstructions with fewer dependencies. Most popularly, we have the E2VID model proposed by Rebecq et al. [5], [6] that trains a recurrent network using pairs of grayscale frames and event streams to generate high-speed videos that can reach more than 5000 frames per second.

Moreover, Scheerlinck et al. [10] work on making the reconstruction process faster and less computationally expensive by applying high-pass event filtering as a preprocessing step before reconstruction. Interestingly, Paredes-Valles et al. [11] introduced the first self-supervised reconstruction model by leveraging the inner workings of event cameras and relying on their circuit functionality and optical flow to train without ground-truth data. Zhang et al. [12] further advance self-supervised event-based reconstruction by targeting blurry RGB videos and overcoming their limitations through motion deblurring and frame interpolation to produce clear, high-quality frames. Another significant approach is the combination of CNN with transformers, where Fox et al. [13] propose an unsupervised reconstruction model that assigns confidence weights to events and uses interpolation to improve brightness reconstruction without relying on training data. Despite their advancements, these reconstruction algorithms overlook security and privacy concerns, motivating our proposal of a reconstruction algorithm that integrates privacy-preserving principles by design.

B. Event Data Privacy

As event cameras play a growing role in surveillance and biometrics, ensuring privacy in event streams is crucial. Advances in reconstruction enable adversaries to extract identifiable information, challenging their privacy-preserving reputation. In the event domain, Du et al. [14] were the pioneers in examining the security of dynamic vision sensors and developed a framework to scramble the positions of events. However, this only protected event streams during transmission and storage and still required decrypting for utilization. This motivated Ahmad et al. [7] to develop an anonymization algorithm based on auto-encoders that learns to anonymize event streams in a way that distorts reconstructed frames but still maintains performance on the task of person re-identification. In contrast, our approach directly operates on reconstructed frames, preserving their utility for human-centered downstream tasks while anonymizing facial information to avoid re-identification.

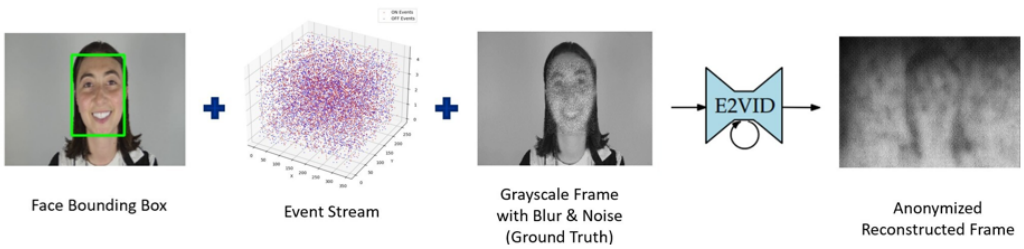
Focusing on face applications, Zhang et al. [4] recently proposed a new unsupervised approach towards encryption by introducing well-crafted noise into raw events until they are scrambled to counter inversion attacks. Their method was tested against these attacks and proved to conceal facial identities. The latter is similar to what we are trying to do by introducing the concept of noise, however, they introduce noise at the level of the event stream to distort reconstructed frames, while we introduce noise at the level of the reconstruction model to detect and generate distorted faces in frames by design.

III. Methodology

A. Problem formulation

Event cameras excel in capturing motion without blur and maintaining performance in challenging lighting conditions, offering a distinct advantage over traditional frame-based systems. Paradoxically, their precise spatiotemporal data nature also renders them vulnerable to privacy attacks aimed at extracting identifiable features, especially since they are often deployed in sensitive or high-security areas. Given that, our goal is to reconstruct intensity images or videos from such

(a) The original E2VID framework.



(b) Our proposed E2PRIV framework.

Fig. 1: Comparison between the original E2VID framework (a) and our proposed E2PRIV framework (b). Both frameworks are visualized on the same dataset (VETEX) for consistency; however, E2VID was originally trained on a different dataset.

event data while anonymizing sensitive facial regions and preserving background details and motion cues—achieving privacy protection without degrading downstream tasks like action recognition.

In this work, we focus on anonymization from a machine perspective rather than a human-centered one. Due to the low spatial resolution of the reconstructed frames, humans perceive them as low quality or degraded, especially when we use our anonymization model as seen in Figs. 2 and 3. However, for machine-based analysis, the crucial information lies in the motion edges and spatiotemporal patterns preserved in the rapid event updates. This highlights the trade-off between human visual quality and machine interpretability, as humans rely on fine details and textures, while machine learning models extract structural patterns, making anonymized frames visually degraded yet still functional for deep learning tasks.

B. Event Data Representation

As mentioned before, event cameras produce a sparse and asynchronous representation of motion in a scene. Each event e_k is defined by its timestamp t_k, pixel coordinates (x_k,y_k), and polarity p_k, indicating whether the brightness at a pixel increased or decreased. These events are triggered at pixel-level when the change in brightness exceeds a predefined contrast threshold, ensuring that only significant changes in a scene are recorded.

For our model, we adopt the same representation used for training the E2VID model, where event data is represented as a spatiotemporal voxel grid. Using this approach, the continuous event stream is discretized into a fixed number of temporal bins, allowing the data to be structured into a tensor suitable for processing with recurrent neural networks (RNNs).

C. Datasets

In this work, we use two datasets, each serving a distinct purpose. The VETEX dataset [15] is ideal for training the reconstruction model as it provides paired RGB and real event data for close-up faces. For evaluation, we use both the VETEX and the FES dataset [16], which contains continuous event streams of faces captured under various conditions. The datasets are described as follows:

VETEX dataset: This trimodal dataset (RGB, Event, Thermal) is made up of almost 840 videos per modality. It was initially collected for the purpose of microexpression recognition with 7 classes, however, for our work, we use it regardless of the class as it provides event data of faces moving in front of the camera along with RGB pairs. For the purpose of face anonymization, we utilize only the RGB and event modalities, which provide around 2.75 hours of data each. It also features a diverse distribution across gender (15 males, 5 females), nationality (10 nationalities), occlusions (10 participants wearing glasses), and lighting conditions (studio and ambient light).

FES dataset: This dataset is a benchmark dataset for faces in event streams, originally collected for the purpose of face detection. It contains 689 minutes of data collected from 73 subjects, balanced between male and female, in both controlled and wild environments. Additionally, each face is annotated with bounding boxes and five facial landmarks, making it well-suited for face detection tasks.

D. Training Data

Our training data is made up of sets of three key inputs: grayscale frames serving as reconstruction ground-truths, event data streams as primary inputs, and their corresponding face bounding box annotations. However, a key challenge in preparing this dataset is ensuring precise synchronization between the RGB frames and the event streams. To address this, we synthesize event streams from the RGB videos of the VETEX dataset using the V2E simulator [17] with noisy configuration. We ensure accurate alignment by extracting RGB frames at precise intervals of 33.33 ms to match the dvs_exposure parameter of 0.033 used during the simulation. Finally, we use the DeepFace face detector [18] with the RetinaFace backend to detect faces in the RGB frames and generate bounding boxes, which are then mapped onto the corresponding event streams.

E. Network Architecture and Training

Our proposed framework is summarized in Fig. 1(b), where we apply transfer learning on the E2VID model, which is built on a recurrent, fully convolutional U-Net architecture. Fig. 1(a) also presents the original E2VID framework for comparison, highlighting the improvements we introduce. We select E2VID as our baseline reconstruction model not only due to its widespread adoption in state-of-the-art research but also for its robust handling of event-based artifacts that occur at the boundaries of moving objects, allowing it to generate high-speed reconstructed images under diverse conditions.

Moreover, to adapt E2VID for face anonymization, we train using the VETEX dataset, which provides RGB-synthetic event pairs and their corresponding face bounding boxes. This transfer learning approach ensures that the model benefits from the robust feature extraction capabilities of the pre-trained weights, adapting them to the task of face anonymization in event-based reconstruction. Our training process introduces two key contributions to the standard E2VID reconstruction pipeline:

Dynamic Blurring with Noise Injection: Inspired by the approach proposed by Yuan et al [19], which demonstrates the effectiveness of hybrid obfuscation techniques in preserving privacy, we combine blurring and noise injection during training to enhance facial anonymization. For dynamic blurring, we introduce a random degree of blur to the face region in each target frame, as defined by the bounding boxes, forcing the model to learn to conceal facial features under varying conditions. Additionally, we inject Gaussian noise into the blurred regions, providing a simpler feature for the model to learn alongside the more complex blur. This choice of noise is supported by prior work [20], which has shown that adding noise effectively obfuscates sensitive facial information, reinforcing its role as an additional privacy-preserving mechanism in our approach.

Double Loss Framework: We use separate losses for the face and background regions. For the face region, we combine the Learned Perceptual Image Patch Similarity (LPIPS) loss, originally used by Rebecq et al. for training E2VID, with L1 regularization loss. LPIPS ensures perceptual similarity by capturing high-level features, while L1 adds pixel-level consistency, enabling effective anonymization. For the background region, we use the Structural Similarity (SSIM) loss to preserve structural details and ensure that the reconstructed non-facial areas remain visually consistent with the input.

F. Training Implementation

For our face anonymization approach, we split the synthetic VETEX dataset into training and testing sets using an 80-20 ratio. We also we use the original E2VID code implementation [5] and initialize training with the pre-trained E2VID weights trained on the MS-COCO dataset.

During training, we sample 1 in every 10 consecutive frames to avoid redundancy in the input data. Moreover, we apply dynamic Gaussian blur with random kernel size ranging from 5 to 30 and add Gaussian noise with 0 mean and 0.05 standard deviation using PyTorch implementation. We train the model for 3 epochs using ADAM optimizer with a learning rate of 0.0001 and batch size of 8.

IV. Experimental Results

In this section, we evaluate our proposed framework by assessing its performance on the VETEX and FES datasets. For testing, we use real event streams from the VETEX dataset, selecting one video per person from each class (140 videos in total) to evaluate generalization. Additionally, we test on a subset of the FES dataset, selecting 20 ten-second videos with varying face angles and distances from the camera. This evaluation allows us to demonstrate the impact of our approach by comparing the performance of E2PRIV and the original E2VID.

A. Evaluation Protocol

Our goal is to anonymize faces in event streams, ensuring that face detection models cannot identify a face in the reconstructed frames. More strictly, for face verification, we aim to prevent re-identification, ensuring that even if a face is detected, it cannot be matched to the same individual before and after anonymization. We measure success by a significant drop in detection rates and recognition confidence compared to E2VID-reconstructed frames, indicating effective anonymization. For clarity, we sometimes refer to the set of reconstructed frames from an event stream as a reconstructed video in this section.

Beyond quantitative metrics, we assess reconstructions through human visual inspection to evaluate anonymization and background preservation. Fig. 2 shows that VETEX frames appear less detailed due to lower event density—since event cameras depend on motion, fewer events result in weaker reconstructions. In contrast, FES reconstructions appear sharper due to higher motion activity. While these images seem heavily degraded from a human perspective, our goal is to evaluate anonymization from a machine perspective, where recognition relies on spatiotemporal features like edge patterns and temporal event distributions rather than visual clarity.

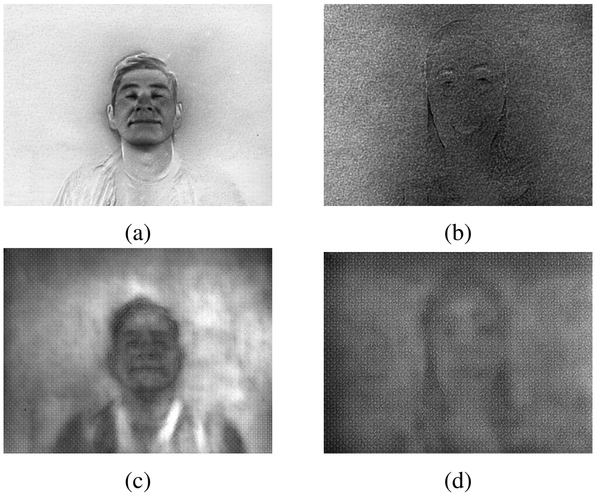


Fig. 2: Examples of reconstructed frames for faces in VETEX dataset (right) and FES dataset (left). Rows represent reconstructions using E2VID (top row) and E2PRIV (bottom row). For visualization purposes, the brightness of the images has been enhanced to improve readability.

B. *Face Detection*

We evaluate our model's anonymization effectiveness by comparing RetinaFace detection rates on frames reconstructed by E2VID and our E2PRIV framework, where an event stream is considered detected if at least one frame within the reconstructed sequence contains a face. Table I shows that for the VETEX dataset, E2PRIV reduces face detection by 6%, given that the original detection rate of this dataset (25.09%) is relatively low compared to that of the FES dataset (89.59%). This proves that, despite being trained on synthetic data, our model generalizes well to real-world event streams in VETEX.

TABLE I: Face Detection Rates in Reconstructed Frames

Model	VETEX dataset	FES dataset
E2VID	25.09%	89.59%
E2PRIV	19%	36.12%

As for the FES dataset, the reduction in detected faces is even more pronounced. E2PRIV achieves a significant 2.48-fold reduction, corresponding to over a 50% decrease in detected faces. A possible explanation for the smaller reduction in face detection for the VETEX dataset compared to the FES dataset is that the model was trained on VETEX, potentially introducing some bias or prior knowledge about the faces in this dataset, making them slightly more recognizable post-anonymization. These results prove the effectiveness of our approach in anonymizing facial information in reconstructed frames.

C. *Face Verification*

To further evaluate the privacy preservation of our model, we assess face verification, determining whether a face can still be matched to the same person before and after anonymization. We generate 32 original-anonymized frame pairs per video from the FES dataset, where the original frames correspond to E2VID reconstructions, and compute verification accuracy using three DeepFace models: VGG-Face, ArcFace, and DeepFace, applying face verification only to faces that were previously detected by the face detector. To establish a baseline, we measure intra-frame verification within E2VID by using randomly paired frames from each video.

TABLE II: Face Verification Results for Reconstructed Frames (FES Dataset). (Avg. Sim. Dist.: Average Similarity Distance).

Model	E2VID		E2PRIV	
	Accuracy	Avg. Sim. Dist	Accuracy	Avg. Sim. Dist
VGG-Face	83.91%	0.48	61%	0.61
ArcFace	89.23%	0.24	88.01%	0.24
DeepFace	20.27%	0.33	12.5%	0.34

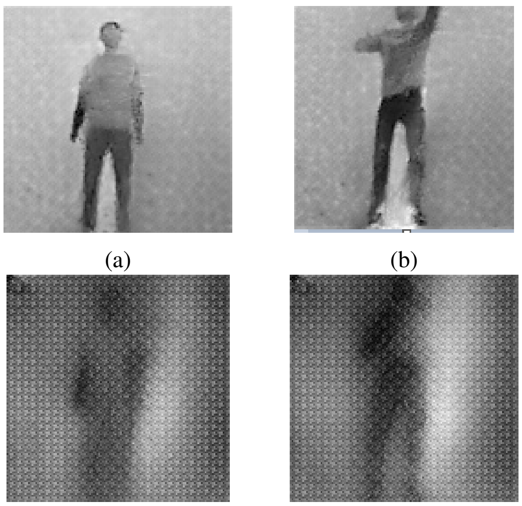
Table II presents the verification accuracy and average similarity distance for the faces in the reconstructed frames. We prove that E2PRIV significantly reduces verification accuracy for DeepFace (12.5%) and VGG-Face (61%) compared to E2VID. However, ArcFace remains highly resilient, maintaining high accuracy (88.01%), suggesting that our approach is insufficient for certain critical applications. It is important to note that these results follow the reduction in detected faces, confirming that E2PRIV effectively disrupts most benchmark verification models, except for ArcFace, which appears resistant, likely due to its training on diverse distortions. Additionally, E2VID-reconstructed faces exhibit lower similarity distances, indicating that they retain more identifiable facial features.

V. Ablation Study: Action Recognition

After demonstrating that our model effectively anonymizes faces in event streams, we study its impact on human-centered applications, particularly action recognition, to ensure it maintains usability while protecting privacy. We evaluate performance on the DailyDVS benchmark dataset, which includes 12 action classes, and compare the action recognition accuracy using frames reconstructed by E2VID and our E2PRIV model. Additionally, we examine the effect of facial presence by creating a test subset using anonymized frames that correspond to those where RetinaFace detected a face in the original E2VID reconstructions. Consequently, we train a C3D PyTorch model for 100 epochs on frames from all models under identical settings.

As shown in Table III, E2PRIV exhibits only a minimal accuracy drop (1.82%) compared to E2VID, confirming that anonymization does not degrade action recognition performance. Interestingly, when trained on face-containing data, E2PRIV achieves nearly the same accuracy as E2VID (95.68% vs. 95.85%), suggesting that facial motion cues—though anonymized—still contribute to action classification.

Fig. 3 presents a visual comparison of reconstructed frames from the DailyDVS dataset, illustrating the effect of anonymization on action sequences. Despite appearing heavily degraded to humans, the reconstructed frames retain the essential motion information required for action recognition, achieving comparable results while significantly reducing face detection and recognition rates.



(a) (b) (c) (d)

Fig. 3: Examples of reconstructed frames from DailyDVS dataset for a person jumping between two positions. Rows represent reconstructions using E2VID (top row) and E2PRIV (bottom row).

TABLE III: Action Recognition Accuracy on DailyDVS dataset.

Model	Accuracy in %
E2VID	95.85%
E2PRIV (frames with detected faces)	95.68%
E2PRIV	94.03%

VI. Conclusion

In the context of event cameras, research predominantly frames event reconstruction as either a tool or a potential attack, focusing on either achieving high-quality event-to-video reconstruction or developing privacy-preserving algorithms to anonymize event streams. To bridge this gap, we proposed E2PRIV, a secure event-to-video reconstruction framework that integrates privacy-preserving principles to anonymize sensitive facial features while maintaining the utility of the reconstructed frames for downstream tasks. We demonstrated that our model effectively obscures facial features, as shown by significant reductions in face detection and recognition rates. At the same time, E2PRIV preserves performance in action recognition tasks, proving its suitability as a privacy-aware substitute for traditional reconstruction methods. While E2PRIV achieves robust performance, future work could explore extending the approach to anonymize other biometric features, such as gait or body shape, or integrating adaptive anonymization techniques tailored to specific application requirements.

Acknowledgment

This research is a part of the HEIMDALL project, funded by the BPI as part of the AAP I-Démo.

References

[1] S. S. Prasad, N. K. Mehta, H. Kumar, A. Banerjee, S. Saurav, and S. Singh, "Hybrid snn-based privacy-preserving fall detection using neuromorphic sensors," in *Proc. Fourteenth Indian Conf. Comput. Vis., Graphics, Image Process. (ICVGIP)*, Dec. 2023, pp. 1–9.

[2] Y. Dong, Y. Li, D. Zhao, G. Shen, and Y. Zeng, "Bullying10k: A large-scale neuromorphic dataset towards privacy-preserving bullying recognition," *Adv. Neural Inf. Process. Syst. (NeurIPS)*, vol. 36, 2024.

[3] S. Kumawat and H. Nagahara, "Privacy-preserving action recognition via motion difference quantization," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*. Springer Nature Switzerland, Oct. 2022, pp. 518–534.

[4] P. Zhang, S. Zhu, and E. Y. Lam, "Event encryption: Rethinking privacy exposure for neuromorphic imaging," *Neuromorphic Comput. Eng.*, vol. 4, no. 1, p. 014002, 2024.

[5] H. Rebecq, R. Ranftl, V. Koltun, and D. Scaramuzza, "High speed and high dynamic range video with an event camera," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 6, pp. 1964–1980, Jun. 2021.

[6] ——, "Events-to-video: Bringing modern computer vision to event cameras," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2019, pp. 3857–3866.

[7] S. Ahmad, P. Morerio, and A. D. Bue, "Person re-identification without identification via event anonymization," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, 2023, pp. 11 132–11 141.

[8] M. Cook, L. Gugelmann, F. Jug, C. Krautz, and A. Steger, "Interacting maps for fast visual interpretation," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, 2011, pp. 770–776.

[9] H. Kim, A. Handa, R. Benosman, S. Ieng, and A. Davison, "Simultaneous mosaicing and tracking with an event camera," in *Proc. British Mach. Vis. Conf. (BMVC)*, 2014.

[10] C. Scheerlinck, N. Barnes, and R. Mahony, "Continuous-time intensity estimation using event cameras," in *Proc. Asian Conf. Comput. Vis. (ACCV)*, Dec. 2018, pp. 308–324.

[11] F. Paredes-Vallés and G. C. D. Croon, "Back to event basics: Self-supervised learning of image reconstruction for event cameras via photometric constancy," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 3446–3455.

[12] X. Zhang and L. Yu, "Unifying motion deblurring and frame interpolation with events," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2022, pp. 17 765–17 774.

[13] G. Fox, X. Pan, A. Tewari, M. Elgharib, and C. Theobalt, "Unsupervised event-based video reconstruction," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV)*, 2024, pp. 4179–4188.

[14] B. Du *et al.*, "Event encryption for neuromorphic vision sensors: Framework, algorithm, and evaluation," *Sensors*, vol. 21, no. 13, p. 4320, 2021.

[15] M. Adra, N. Mirabet-Herranz, and J. L. Dugelay, "Beyond rgb: Tri-modal microexpression recognition with rgb, thermal, and event data," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, 2024.

[16] U. Bissarinova, T. Rakhimzhanova, D. Kenzhebalin, and H. A. Varol, "Faces in event streams (fes): An annotated face dataset for event cameras," *Authorea Preprints*, 2023.

[17] Y. Hu, S. C. Liu, and T. Delbruck, "v2e: From video frames to realistic dvs events," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, 2021.

[18] S. I. Serengil and A. Ozpinar, "Lightface: A hybrid deep face recognition framework," in *2020 Innovations in Intelligent Systems and Applications Conference (ASYU)*. IEEE, 2020, pp. 23–27. [Online]. Available: https://ieeexplore.ieee.org/document/9259802

[19] L. Yuan, W. Chen, X. Pu, Y. Zhang, H. Li, Y. Zhang, X. Gao, and T. Ebrahimi, "Pro-face c: Privacy-preserving recognition of obfuscated face via feature compensation," *IEEE Transactions on Information Forensics and Security*, 2024.

[20] T. Zhang, Z. He, and R. B. Lee, "Privacy-preserving machine learning through data obfuscation," *arXiv preprint arXiv:1807.01860*, 2018.