Event-based Visual Deformation Measurement

Yuliang Wu[1], Wei Zhai[1,†], Yuxin Cui[1], Tiesong Zhao[2], Yang Cao[1], Zheng-Jun Zha[1]

[1]MoE Key Laboratory of Brain-inspired Intelligent Perception and Cognition,
University of Science and Technology of China [2]Fuzhou University

{tronliang@mail., wzhai056@, yuxincui@mail.}ustc.edu.cn
t.zhao@fzu.edu.cn {forrest@, zhazj@}ustc.edu.cn

Abstract

Visual Deformation Measurement (VDM) aims to recover dense deformation fields by tracking surface motion from camera observations. Traditional image-based methods rely on minimal inter-frame motion to constrain the correspondence search space, which limits their applicability to highly dynamic scenes or necessitates high-speed cameras at the cost of prohibitive storage and computational overhead. We propose an event-frame fusion framework that exploits events for temporally dense motion cues and frames for spatially dense precise estimation. Revisiting the solid elastic modeling prior, we propose an Affine Invariant Simplicial (AIS) framework. It partitions the deformation field into linearized sub-regions with low-parametric representation, effectively mitigating motion ambiguities arising from sparse and noisy events. To speed up optimization and reduce error accumulation, a neighborhood-greedy optimization strategy is introduced, enabling well-converged sub-regions to guide their poorly-converged neighbors, effectively suppress local error accumulation in long-term dense tracking. To evaluate the proposed method, a benchmark dataset with temporally aligned event streams and frames is established, encompassing over 120 sequences spanning diverse deformation scenarios. Experimental results show that our method outperforms the state-of-the-art baseline by 1.6× in survival rate. Remarkably, it achieves this using only 18.9% of the data storage and processing resources of high-speed video methods. Project page: https://wyl-ovo.github.io/EVDM/.

1. Introduction

Deformation is ubiquitous in the real world, ranging from subtle material strain to large-scale structural changes. Visual Deformation Measurement (VDM) recovers deformation fields by accurately tracking surface motion from camera observations, serving as an efficient non-contact technique for structural monitoring [29, 38], mechanics analysis [6, 26, 30], and biomechanics research [54].

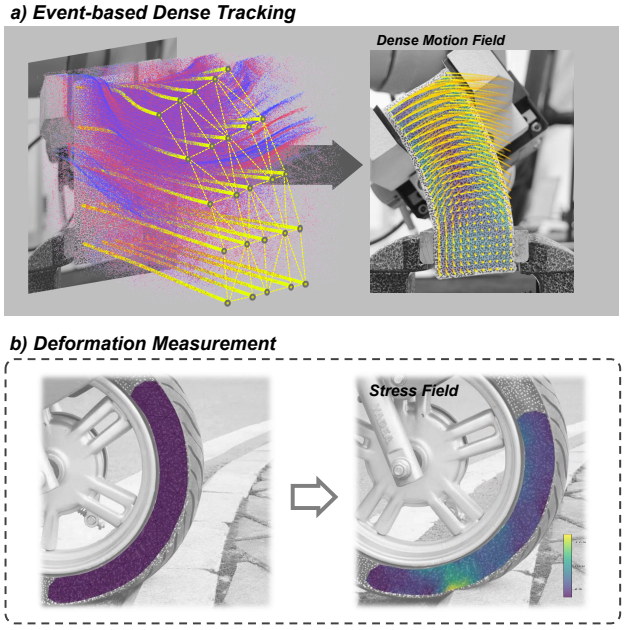

Figure 1. Overview of Event-based Deformation Measurement. (a) Our approach regresses deformation fields by leveraging affine-invariant spatiotemporal trajectory flows extracted from both events and images. (b) Measuring tire deformation while rolling over a step obstacle. Demonstrating robust and accurate measurement capabilities of our event-based VDM approach for objects experiencing rapid self-motion.

As a dense tracking vision problem, the goal of VDM is to estimate the surface motion of a target at the current moment relative to its initial undeformed state. However, unlike rigid motion with only 6-DoF, deformable surfaces have significantly higher degrees of freedom where surface points can move with considerable independence.

†Corresponding author.

This poses fundamental challenges to image-based methods stemming from: (i) an intractably large correspondence search space, and (ii) texture similarity and geometric changes from deformation that impede reliable feature matching. To mitigate these challenges, existing image-based deformation measurement methods rely on costly high-speed cameras to capture videos with minimal inter-frame motion, thereby constraining the correspondence search space. However, this strategy incurs prohibitive storage and computational costs from processing massive redundant frames, limiting the practicality and scalability.

In this paper, we address these challenges by introducing an event-frame hybrid system that exploits the complementary nature of two vision modalities. Events, being temporally dense yet spatially sparse, capture motion dynamics efficiently without the massive redundancy. Conversely, conventional frames offer spatially dense, low-noise observations to ensure precise dense field estimation. This synergy enables accurate deformation tracking without the prohibitive costs of traditional high-speed imaging. However, employing the hybrid system for deformation measurement presents numerous challenges. Firstly, the sparse spatiotemporal sampling inherent to event cameras, coupled with their high noise levels, introduces ambiguities in the estimation of dense field motion information.

Secondly, long-term tracking suffers from error accumulation, which is significantly amplified in dense tracking due to the large number of descriptors at each time step. The accumulated local errors can potentially lead to global failure.

In this paper, we propose an Affine Invariant Simplicial (AIS) framework to locally linearize the deformation field, reduce the number of motion field descriptors and mitigate motion estimation ambiguities. Meanwhile, we introduce a neighborhood-greedy optimization strategy that corrects mismatched sub-regions using their well-converged neighbors, thereby reducing error accumulation in dense tracking and achieving long term global optimality.

In the proposed AIS framework, the nonlinear deformation field of the object surface is systematically divided into multiple sub-regions. Within each sub-region, deformation parameters are linearized using simplicial motion parameterization, effectively reducing their dimensionality. Then, leveraging event contrast maximization and image cross-correlation as optimization objectives, a coarse-to-fine scheme is adopted to hierarchically decouple and sequentially solve for: rigid self-motion and fine-grained deformations across multi-level mesh subdivisions. The AIS framework hits two birds with one stone: it reduces the difficulty of solving the high-dimensional deformation field and addresses the ambiguity in motion information estimation from event data. Building upon the framework, we employ the neighborhood-greedy optimization strategy. We design a global convergence criterion to distinguish poorly-converged subregions that significantly deviate from mean convergence states. Then, Leveraging continuity priors, the optimization of this subregions are guided by their well-converged counterparts, significantly reducing long-term error accumulation in dense tracking.

To establish a comprehensive evaluation benchmark for our method and a series of baselines, we captured a dataset comprising over 120 real-world sequences spanning various deformation modes and magnitudes, with temporally aligned event and frame data. Ground truth deformation fields derived from high-speed videos are provided to facilitate quantitative evaluation and advance research.

In summary, our contributions are as follows.

1. A hybrid VDM system that leverages the high temporal resolution of event cameras to capture rapid texture motion, enabling accurate measurement of large-scale deformations and those involving significant self-motion.
2. An affine invariant simplicial framework is proposed for robust deformation measurement from events and frames. This formulation retains the expressive power for complex deformation fields while mitigating the motion estimation ambiguity and noise sensitivity inherent to spacial-temporal sparse event data.
3. A neighborhood-greedy optimization strategy is proposed to enable well-converged sub-regions to guide their poorly-converged neighbors, mitigating local error accumulation in long term dense tracking.
4. Extensive experiments demonstrate the competitiveness of the proposed approach, achieving a 1.6× survival rate of SOTA baseline, while maintaining superior EPE accuracy with only 18.9% of the data storage and processing cost of traditional high-speed video methods.

2. Related Work

2.1. Visual Deformation Measurement

As an efficient non-contact deformation sensing approach, visual deformation measurement (VDM) has been widely applied in scientific research [15, 22, 27, 36, 44, 54, 57] and engineering fields [6, 26]. Image-based VDM algorithms use correlation criteria to match intensity values between regions before and after deformation,

$$C = \sum_{i=1}^{N} \left(af\left(x_i, y_i\right) + b - g\left(x_i', y_i'\right) \right)^2 . \tag{1}$$

Where $(x_i, y_i) \in R$ and $(x_i', y_i') \in R'$ denote corresponding point pairs in the original and deformed regions, respectively. Here, a represents a scale factor and b represents an intensity offset. The task of the VDM is to minimize the coefficient C in Eq. 1 to detect the best matching. However, unlike rigid body motion tracking in everyday scenarios that only requires estimating a limited number of parameters, VDM is fundamentally a dense tracking task that

demands high accuracy, involves complex variations, and requires optimization over a large parameter space, making it a complex high-dimensional optimization problem. To reduce the complexity of subset matching, VDM methods adopt the spatiotemporal continuity prior of motion and search for matching subsets in the vicinity of the original subset positions [20, 32].

This limitation prevents them from effectively handling large deformations and large displacement scenarios with rapid object self-motion [4, 47, 51, 52], or necessitates the use of costly high-speed cameras, which introduces a significant data storage burden, greatly restricting the potential applications of the VDM method.

2.2. Event-based Vision

Different from traditional frame-based camera, the neuromorphic event camera abandons the fixed-interval sampling approach. Instead, it uses trigger-based sampling to passively capture dynamic information from the scene [5, 10, 11, 19, 28, 39, 45, 46]. Specifically, the i-th event $e_i := (x_i, y_i, t_i, p_i)$ is triggered at time t_i whenever the log-scale scene brightness $\mathbf{I}(t) = \log(I_{x_i,y_i}(t))$ change exceeds threshold c, i.e,

$$p_i = \begin{cases} 1, & \text{if } \mathbf{I}(t_i) - \mathbf{I}(t_i - \Delta t_i) \geq c, \\ -1, & \text{if } \mathbf{I}(t_i) - \mathbf{I}(t_i - \Delta t_i) \leq -c, \end{cases} \tag{2}$$

Here, Δt_i indicates the time since the last event at (x_i, y_i), and $p_i \in \{-1, +1\}$ denotes the event polarity. This sampling strategy provides event cameras with high temporal resolution [7, 42, 43], high dynamic range [34, 48, 53, 59], and low storage redundancy [17], making them ideal for continuous motion observation in non-rigid shape estimation and reconstruction tasks [33, 35, 49, 50].

2.3. Contrast Maximization

Currently, there are various model-based [9, 12, 41] and learning-based [13, 14, 40, 58] approaches being capable of motion tracking from event streams at the pixel level. Regrettably, there is currently no event dataset suitable for VDM (Visual Deformation Modeling) tasks. Moreover, creating large-scale deformation datasets is costly and inefficient. In this work, the data-independent contrast maximization (CM) method is employed for motion estimation. As a versatile method, CM works by maximizing the contrast of the image of warped events (IWE) to find the motion field that best fits the event set [9, 14, 41]. This allows it to be more robust than the fixed-template matching methods [12], especially for handling feature shape changes caused by deformations. However, due to spatiotemporal sparsity and high noise level in event data, pixel-level CM-based motion estimation suffers from significant ambiguity. Existing CM frameworks address this through low-parameter modeling for global motion (e.g., camera ego-motion or planar scenes[9]) or smoothness/regularization constraints for local neighborhoods [18]. However, neither approach applies to VDM tasks: deformations lack global parametric models, while smoothness constraints suppress the framework's ability to capture deformation details. Incorporating physical priors, such as solid elastic modeling, offers a promising solution to capture complex deformable behaviors [1, 2, 24, 31]. In this paper, a simplicial parameterization framework is proposed to address the dilemma. The surface region of the deformed object is divided into a set of continuous local sub-regions and simplify the deformation of each sub-region as a low-parameter description. This allows achieving low-ambiguity CM motion estimation while preserving the ability to describe the deformation.

3. Methodology

In this section, we present our approach for dense tracking of deformable objects using hybrid event-frame data. Sec. 3.1 formulates the VDM problem and introduces the affine-invariant property that enables efficient field representation. Sec. 3.2 describes how we decompose the deformation field into sparse anchor trajectories and associate events/images with these anchors through geometric operations. Sec. 3.3 details our coarse-to-fine and neighborhood-greedy optimization strategy for efficient solving.

3.1. Preliminaries

VDM problem. The goal of VDM is to recover the surface mapping relationship before and after object deformation through visual methods, i.e., the deformation field u. Under the Lagrangian perspective, we formulate the problem as follows. At any time t_i, a material point initially located at position $X \in S$ on the object surface S deforms to its current position:

$$x(X, t_i) = X + u(X, t_i), \quad X \in S, \tag{3}$$

where $u(X, t)$ represents the displacement vector of the material point initially located at X on surface S. In traditional image-based tasks, this is defined as a simple mapping between two frames, whereas in event-based VDM, we need to construct a data structure to represent the deformation field in continuous spatiotemporal space.

Affine-invariant property. For any given triangular sub-region defined as: $\sigma = \text{conv}\{X_1, X_2, X_3\}$[1], the affine interpolation operator can be expressed as:

$$I[f](X) = \sum_{i=1}^{3} \lambda_i(X) f(X_i), \quad \text{s.t.} \sum_{i=1}^{3} \lambda_i(X) = 1. \tag{4}$$

For any arbitrary affine function of the form $f(X) = AX + b$, the interpolation operator satisfies the reproduction property $I[f] = f$. This fundamental property enables modeling of spatiotemporal deformation fields via low-parametric sparse triangle sub-regions.

[1]The $conv\{.\}$ operator denotes the convex hull

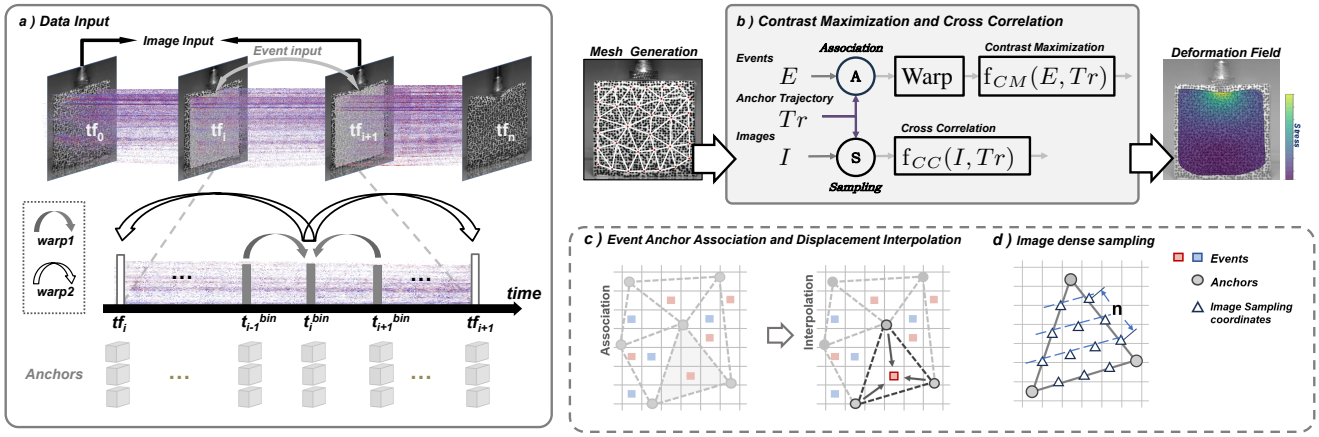

Figure 2. **Illustration of the key steps of our method.** (a) Data input and event warping strategy. (b) Overall pipeline of the proposed event-based VDM framework. (c) Illustration of event-to-subregion association through vector cross product with anchor trajectories and subsequent displacement interpolation. (d) Image intensity sampling scheme within each subregion.

3.2. Dense Tracking with Simplicial Framework

In this paper, we introduce the affine invariant simplicial framework (AIS), which represents the spatiotemporal deformation field as lightweight trajectories $\{\mathrm{Tr}_j(t)\}$. Specifically, high-dimensional nonlinear deformation field u is decomposed into N triangular sub-regions $T_k, k = 1, \ldots, SN$, where the deformed position within each triangle follows an affine transformation:

$$x(X) = X + u(X) = A_k X + b_k, \quad \forall X \in T_k, \tag{5}$$

$A_k \in \mathbb{R}^{2\times 2}$ is the local deformation gradient matrix and $b_k \in \mathbb{R}^2$ is the translation vector. Larger SN improves nonlinear expressiveness but complicates optimization, while smaller SN offers simplicity with reduced none-linear representation ability. Stems from the eq. 4 property, the affine transformation of each sub-region can be losslessly described by the motion of its vertices, we optimize the trajectories of these anchor points $\{\mathrm{Tr}_j(t)\}$ using events and images to obtain the continuous spatiotemporal field $u(t)$.

Event anchor association and displacement interpolation. During optimization process, the association between individual events and their corresponding trajectories is unknown, and efficiently establishing this association is a key challenge. Prior work [18] employs knn search to find the top-k nearest anchors for each event, then computes displacements via averaging. However, per-event KNN search is time-consuming, and averaging-based interpolation introduce significant errors in deformation measurement. In this work, we perform same-side test event association and using the affine invariant interpolation to address this issue, as shown in Fig. 2 (d. Consider the spatiotemporal motion of an individual sub-region $\sigma(t) = conv\{\mathrm{Tr}_1(t), \mathrm{Tr}_2(t), \mathrm{Tr}_3(t)\}$.

The same-side Test association is performed as follows. Given event coordinate $\mathbf{X}_e = [x_i, y_i]^T$ and triangle vertices at triggering time $\mathrm{Tr}_j(t_i) = [x_j^t, y_j^t]^T$, we compute:

$$C_i = \begin{vmatrix} x^t_{(j \bmod 3)+1} - x^t_j & y^t_{(j \bmod 3)+1} - y^t_j \\ x_i - x^t_j & y_i - y^t_j \end{vmatrix}, \tag{6}$$

where $j \in \{1, 2, 3\}$, and the event is inside if:

$$\mathbf{X}_e \in \sigma(t) \Longleftrightarrow \mathrm{sign}(C_1) = \mathrm{sign}(C_2) = \mathrm{sign}(C_3) \tag{7}$$

Once associated with a subregion, the barycentric interpolation weights λ_i are computed by solving:

$$\mathbf{X}_e = \sum_{k=1}^{3} \lambda_k \mathrm{Tr}_j(t_i), \quad \text{s.t.} \sum_{k=1}^{3} \lambda_k = 1, \tag{8}$$

The solving process are provided in appendix. For any selected timestamp t_{ref}, the event can be warped according to $\mathbf{X}'_e = [x'_i, y'_i]^T = \sum_{k=1}^{3} \lambda_k \mathrm{Tr}_j(t_{ref})$ to form the image of warped events (IWE). Following the formulation in [16, 37], the IWE is constructed as:

$$\begin{gathered} T_p(\boldsymbol{x}, t_{\text{ref}}) = \frac{\sum_j \kappa(x - x'_j)\kappa(y - y'_j) w(t_j)}{\sum_j \kappa(x - x'_j)\kappa(y - y'_j) + \epsilon}, \\ w(t_j) = 1 - \frac{|t_{ref} - t_j|}{\max_i |t_{ref} - t_i|}, \\ \kappa(a) = \max(0, 1 - |a|), \\ j = \{i \mid p_i = p\}, \quad p' \in \{+1, -1\}, \quad \epsilon \approx 0, \end{gathered} \tag{9}$$

And the contrast maximization objective is formulated as:

$$\mathrm{f}_{\mathrm{CM}}(t_{\text{ref}}) = \frac{\sum_{\boldsymbol{x}} T_{+1}(\boldsymbol{x}, t_{\text{ref}})^2 + T_{-1}(\boldsymbol{x}, t_{\text{ref}})^2}{\sum_{\boldsymbol{x}} [n(\boldsymbol{x}') > 0] + \epsilon}, \tag{10}$$

where $n(\boldsymbol{x}')$ denotes a per-pixel event count of the IWE. To recover anchor motion trajectories, we warp the event

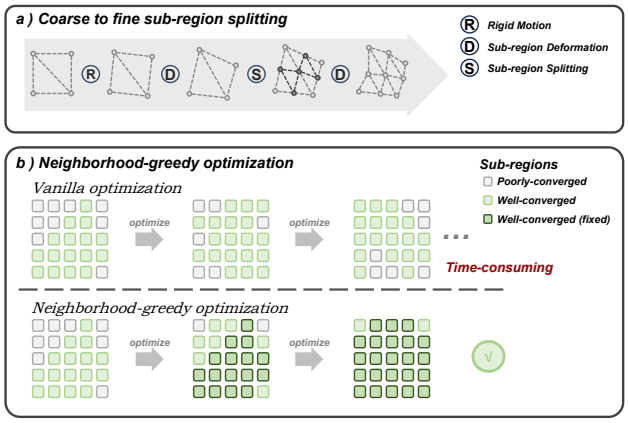

Figure 3. The optimization steps. (a) Hierarchical subregion optimization: we first estimate rigid motion parameters of the object, then progressively split and optimize subregions from coarse to fine. (b) Neighborhood-greedy optimization: well-optimized subregions serve as anchors to guide neighboring regions toward better convergence, achieving faster convergence to the global optimum compared to direct high dimensional optimization.

stream to a set of reference timestamps t_{ref} and jointly optimize the motion parameters $\mathrm{Tr}_j(t_{ref})$ by minimizing the contrast maximization objective.

Image intensity sampling from anchors. Due to the loss of precise grayscale information in sparse events, frames are introduced to refine the deformation field. Specifically, S_1 and S_2 denote the sets of intensity samples at corresponding coordinates across two different frames I_1, I_2. The coordinates are obtained through uniform barycentric sampling within sub-region $\sigma = conv\{\mathrm{Tr}_1, \mathrm{Tr}_2, \mathrm{Tr}_3\}$ at frame capturing time:

$$\mathcal{B} = \left\{ \left(\frac{i}{n}, \frac{j}{n}, 1 - \frac{i+j}{n} \right) \mid i, j \in \mathbb{N}, i+j \leq n \right\}, \tag{11}$$

where n is the number of samples per sub-region edge, and the sampling coordinates are given by $u\,\mathrm{Tr}_1 + v\,\mathrm{Tr}_2 + w\,\mathrm{Tr}_3$ for each $(u, v, w) \in \mathcal{B}$. Then the intensity at each sampling coordinate are obtained from image pixels via bilinear interpolation, and zero-mean normalized cross-correlation is employed to compute the cross-correlation between the intensity samples:

$$\mathrm{f_{CC}}\left(S_1, S_2\right) = \frac{\mathrm{Cov}\left(S_1, S_2\right)}{\sigma_{S_1} \cdot \sigma_{S_2}}. \tag{12}$$

In practice, we optimize the $\mathrm{f_{CC}}$ between the current frame I_i and both the previous frame I_{i-1} and the initial frame I_0.

3.3. Steps of the Optimization

The optimization pipeline iteratively estimates motion within each time window tw in a forward manner. In practice, the events $\mathcal{E}_{k,k+1}$ between consecutive frames $\{f_k, f_{k+1}\}$ are partitioned into M non-overlapping bins with equal numbers of events. Under the assumption that motion within each bin is linear, we optimize trajectories at $M+1$ timestamps.

$$\{\mathrm{Tr}_j(t_i^{bin})\}, i = 0, \ldots, M. \tag{13}$$

Each event undergoes two warping operations to generate IWE for contrast maximization calculation, as shown in Fig.2 a): **Warp 1**: For each time stamps t_i^{bin}, events from neighboring bins are warped to produce an M-channel IWE sequence, capturing short-term motion. **Warp 2**: All events within tw are warped to t_{f_i} and $t_{f_{i+1}}$, generating a 2-channel IWE to ensure global motion continuity. The optimization pipeline follows a coarse-to-fine paradigm, as shown in Fig. 3 a). It first estimates the rigid motion parameters of the target region, then performs initial deformation estimation on coarse subregions, and finally refines the deformation field through iterative splitting of these subregions. The subdivision process is as follows:

$$\begin{aligned} &\mathrm{conv}\{\mathbf{Tr}_1, \mathbf{Tr}_2, \mathbf{Tr}_3\} \rightarrow \\ &\bigcup_{i=1}^{3} \mathrm{conv}\{\mathbf{Tr}_i, \mathbf{M}_i, \mathbf{M}_{i-1}\} \cup \mathrm{conv}\{\mathbf{M}_1, \mathbf{M}_2, \mathbf{M}_3\}, \\ &M_i = \frac{\mathbf{Tr}_i + \mathbf{Tr}_{(i \bmod 3)+1}}{2}, \quad i = 1, 2, 3. \end{aligned} \tag{14}$$

The optimization objective is:

$$\mathrm{f}_{total} = \lambda_1\, \mathrm{f}_{CM}^{Warp1} + \lambda_1\, \mathrm{f}_{CM}^{Warp2} + \lambda_2\, \mathrm{f}_{CC}, \tag{15}$$

where λ_i are iteration-dependent coefficients. This strategy achieves large-scale displacement estimation with low-parameters (rigid motion and coarse subregions), providing reliable initialization for high-parameter, fine-scale deformation estimation.

Neighborhood-greedy optimization strategy. While this pipeline achieves satisfactory convergence for most subregions, long-term tracking involves multiple iterations where errors from a small fraction of unmatched subregions can accumulate and lead to tracking failure. Therefore, it is crucial to pursue a global optimum at each time step to ensure robust and accurate tracking performance. However, jointly optimizing all sub-regions poses significant computational challenges due to the high-dimensional nature of the deformation parameters, making direct global optimization prohibitively time-consuming. To address this issue, we propose a neighborhood-greedy optimization strategy. The key idea is to leverage the continuity prior of deformation fields, using well-converged sub-regions to guide the optimization of their poorly-converged neighbors, and greedily approaching the global optimum. First, the convergence quality of each subregion is assessed by computing the proportion of sampled pixels whose squared error (SE) deviates from the mean squared error (MSE) within the subregion Ω_j:

$$P_j = \frac{1}{N}\sum_{i=1}^{N} 1(\mathrm{SE}(i) > k \cdot \mathrm{MSE})$$
$$\text{converged}\,(\Omega_j) = \begin{cases} \text{False}\,, & \text{if } P_j > \tau \\ \text{True}\,, & \text{otherwise} \end{cases} \tag{16}$$

where k and τ are hyperparameters. Then, anchor points associated with well-converged sub-regions are fixed, and strain continuity constraints f_S are added to the objective function for further optimization: $\mathrm{f}_S = \frac{1}{|E_T|}\sum_{(i,j)\in E_T} \|\mathbf{S}_i - \mathbf{S}_j\|_F^2$, where $\mathbf{S}_i \in \mathbb{R}$ is the von-mises strain at anchor i (detailed computation in supplementary), and E_T denotes the set of edges of the sub-region. As shown in Fig. 3 b), after each optimization, the convergence status is re-evaluated and fixed anchors are greedily accumulated without release, until the global optimum is reached. The following experiments 3 show that the strategy both improves the long-term dense tracking survival rate and reduces the convergence time of the optimization.

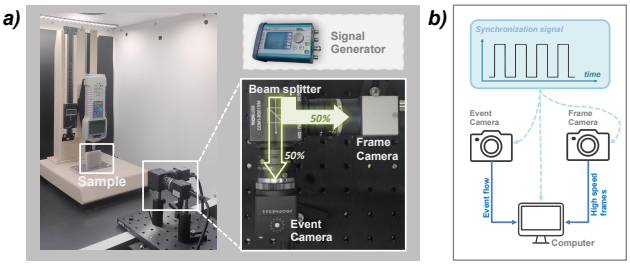

Figure 4. **Hardware configuration of the hybrid system. (a)** Left: Measurement setup where a pressure/stretching machine applies controlled force to the rubber sample. Right: The hybrid system comprises an event camera (Prophesee EVK4) and a grayscale camera. A signal generator produces square wave synchronization signals. The cameras are collocated by mounting a 50:50 split ratio beam splitter in front of them. Spatial calibration is performed before each data acquisition. **(b)** Data transmission topology of the system. A 210Hz square wave signal is used for triggering the frame camera and synchronizing the event stream. Simultaneously acquired event and frame data are transmitted for processing.

4. Experiments

4.1. Event-based VDM Benchmark

Existing real-world VDM datasets are scarce and lack corresponding event data. To comprehensively evaluate our method, we collect a dataset using our hybrid system (Fig. 4), which comprises temporally aligned event streams and high-frame-rate videos (210 fps). The dataset encompasses over 120 diverse deformation motions sequences spanning squeezing, stretching, bending, and cracking, covering diverse scenarios from small-scale (less than 20 pixels) to large-scale displacements (100+ pixels), as well as complex coupled deformation-egomotion dynamics. The ground truth is established by applying VDM algorithms [23] to high-speed video frames, followed by manual refinement.

4.2. Evaluation

Baselines. On the collected dataset, we benchmark our approach against several representative methods: OpenCorr [23], a widely-used open-source model-based VDM algorithm; StrainNet [3], a learning-based VDM method; E-RAFT [13], an event-based optical flow method for short-term motion estimation; and CoTrackerV3 [25], a long-term point tracking method. Given that these methods produce outputs in different formats (e.g., displacement fields, optical flow, point tracks), we adopt a unified evaluation framework to ensure fair comparison. Specifically, a dense set of points is uniformly sampled within the measurement ROI and their displacement $\{u\}$ are tracked across all frames using the respective method. The resulting trajectories are then evaluated against ground-truth annotations. Additionally, to ensure fairness in terms of input information, we evaluate the three image-based methods (OpenCorr [23], StrainNet [3], CoTracker [25]) using high-frame-rate videos generated by TimeLens [42], an interpolation method that synthesizes intermediate frames from the low-frame-rate videos and event streams.

Evaluation metrics. We report three metrics: end-point error (EPE), survival rate, and survival EPE (SEPE). The EPE is computed as: $\mathrm{EPE} = \frac{1}{n}\sum_i \|u(i) - u_{gt}(i)\|_2$, which measures the average accuracy of the optical flow estimation across all points. Following the evaluation protocol in [56], the survival rate measures the average time until dense tracking failure as a fraction of the entire sequence duration, where a tracking failure occurs when the L2 distance of 20% of the sampled points exceeds 5 pixels. Finally, SEPE computes the EPE exclusively over survival points, serving as an accuracy indicator for the deformation field within the convergence region.

Quantitative results. Table 1 presents the experimental results with event and 5 fps frame inputs. For small deformations (5-20 pixels), most methods maintain an EPE below 1.0 and achieve survival rates above 95%. However, as the deformation magnitude increases, all methods exhibit a decline in both tracking accuracy and survival rate. In scenarios with large continuous displacements (100+ pixels), the sota baseline, cotrackerv3, achieves a survival rate of only 45.2%, whereas our method maintains a survival rate of 65.7% with an EPE of 3.204 and SEPE of 0.813, demonstrating robust measurement capability under large motions and displacements. Fig. 5 presents the visualization results of each parallel method. Two representative scenarios are selected: large deformation samples (100+ pixels) under clamping and twisting (I, II), and small deformation samples (5-20 pixels) subjected to tip pressure.

Method	Data input		5-20 pixels			20-100 pixels			100+ pixels		
	Event	Frame	EPE ↓	SEPE ↓	Survival ↑	EPE ↓	SEPE ↓	Survival ↑	EPE ↓	SEPE ↓	Survival ↑
Opencorr [23]		✓	**0.132**	**0.108**	99.1%	1.727	0.478	82.5%	39.931	2.782	25.0%
Opencorr [23] + Timelens [42]	✓	✓	0.227	0.226	**99.5%**	0.819	0.657	89.8%	3.830	1.201	41.3%
StrainNet [3]		✓	1.553	1.146	91.0%	3.413	1.230	61.2%	41.870	1.138	18.8%
StrainNet [3] + Timelens [42]	✓	✓	0.978	0.911	95.2%	1.745	1.603	91.8%	5.189	2.315	35.4%
E-Raft [13]	✓		6.574	1.510	21.7%	21.422	1.632	6.7%	60.312	1.700	2.1%
Cotrackerv3 [25]		✓	0.671	0.671	99.0%	2.138	2.026	91.7%	8.763	2.150	45.2%
Cotrackerv3 [25] + Timelens [42]	✓	✓	0.784	0.783	99.1%	2.387	1.825	87.5%	10.788	2.278	29.1%
Ours	✓	✓	0.155	0.121	99.4%	**0.330**	**0.211**	**92.4%**	**3.204**	**0.813**	**65.7%**

Table 1. **Quantitative comparisons** of our method to the model-based [23] and learning-based [3] VDM algorithms, event-based dense optical flow methods [13], and long-term point tracking algorithms [25]. Image-based methods are enhanced with event-based frame interpolator [42] for thorough comparison. In terms of metrics, lower is better for EPE and SEPE, while higher is better for Survival Rate.

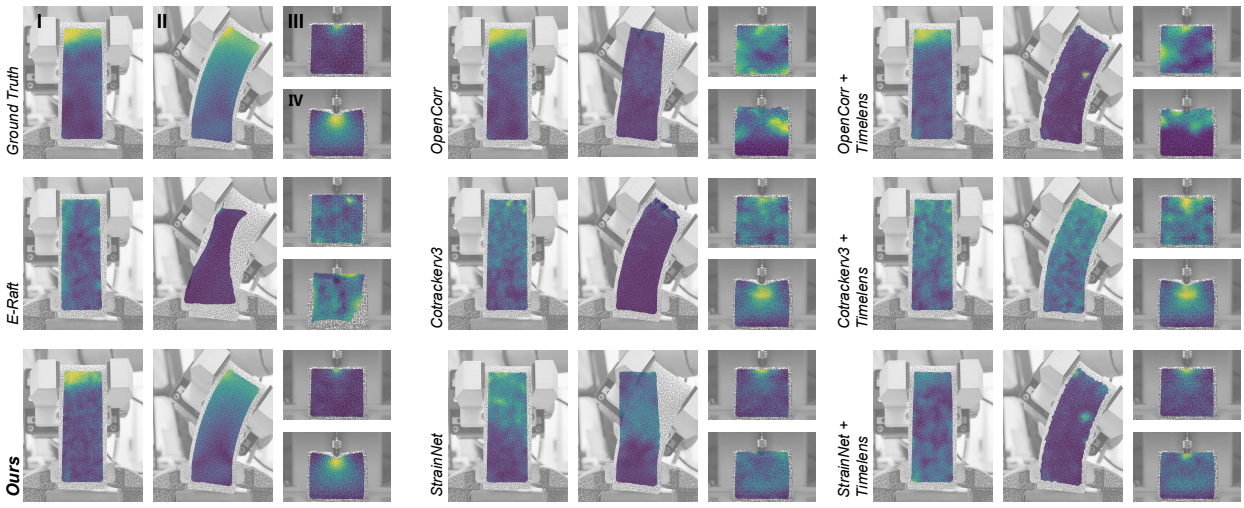

Figure 5. **Qualitative comparisons.** Results on samples under clamping and twisting (I, II), and subjected to tip pressure (III, IV).

4.3. Ablation Study

Data storage burden. We compare the data storage consumption of our method in Table 2. Our method demonstrates significant storage efficiency, requiring only 13% of the storage consumed by 210 fps high-speed cameras, while processing 5 fps frame and event inputs and achieving a competitive survival rate of 65.7% and an SEPE of 0.813.

Displacement interpolation. To validate the effectiveness of our affine-invariant displacement interpolation, we conduct ablation experiments comparing it against four common interpolation methods: (1) Nearest Neighbor, (2) Mean Interpolation, (3) Inverse Distance Weighting, and (4) Gaussian Weighted Interpolation. All experiments are conducted under identical settings except for the interpolation method. As shown in Fig. 7, our method consistently achieves superior deformation measurement accuracy and survival rate compared to non-affine-invariant alternatives.

Optimization strategies. We compare the proposed neighborhood-greedy optimization method with the vanilla optimization method on the entire test set in terms of survival rate and mean convergence time between consecutive frames. As shown in Table 3, the neighborhood-greedy strategy effectively improves the survival rate from 49.0% to 87.1% by reducing error accumulation. Moreover, when incorporating convergence quality assessment for global convergence determination, it achieves a threefold speedup

Data Input			Data Size	EPE ↓	SEPE ↓	Survival ↑
Event	Frame	Frame Fps	(frames + events)			
✓	✓	1	5.6Mb + 78.3Mb	4.710	2.533	32.6%
✓	✓	5	28.1Mb + 78.3Mb	3.204	0.813	65.7%
✓	✓	10	56.3Mb + 78.3Mb	2.110	0.795	67.5%
✓	✓	20	112.6Mb + 78.3Mb	1.618	0.573	71.2%
	✓	5	28.1Mb	39.931	2.782	25.0%
	✓	100	562.5Mb	3.317	0.825	64.3%
	✓	210	1181.3Mb	ground truth calculation		

Table 2. **Comparison of data storage and processing consumption** on the 100+ pixel displacement test set. The upper panel illustrates the performance of our method with event and various frame rate inputs, while the lower shows the performance of the baseline method [23] used for ground truth calculation at different frame rates. The green bars represent our method with 5 fps input, requiring only 18.9% data storage, while the blue bars present the baseline method at 100 fps with comparable performance.

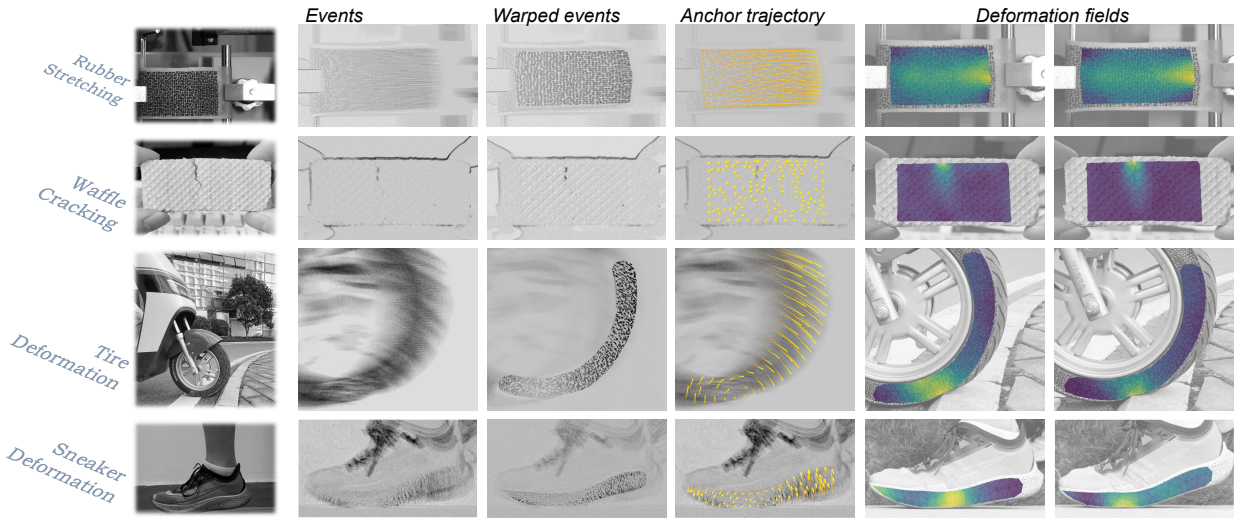

Figure 6. **Results of our system across diverse scenarios.** Each row corresponds to a different deformation scenario. From left to right, each column shows: (i) the scene overview, (ii) event count map before motion estimation, (iii) warped event count map after optimization, (iv) regressed subregion anchor trajectories, and (v) von Mises deformation fields at two time instances during the deformation process.

Strategy	Survival ↑	mean converge time ↓
Neighborhood-greedy optimization	65.7 %	7.2s
Vanilla optimization	39.0 %	26.5s

Table 3. **Ablation studies on optimization strategies.** Experiments are conducted using the multi-scale search combined with the Adam optimizer within the PyTorch framework, utilizing a single NVIDIA RTX 4090 GPU (24GB).

in mean convergence time (from 26.5s to 7.2s).

4.4. Measurement Results

We evaluated our system across multiple representative scenarios to demonstrate the capabilities and application of event-based VDM (see Fig. 6), including: (a) **Rubber band stretching**, which exhibits large-scale elastic deformation that poses significant challenges to VDM methods; (b) **Wafer cracking**, featuring abrupt fracture and discontinuous motion at crack sites; (c) **Tire deformation**, relevant to vehicle safety assessment and predictive maintenance; and (d) **Sneaker deformation**, applicable to gait analysis and athletic performance evaluation. These examples demonstrate the broad applicability of our method from material testing to real-world monitoring applications.

5. Conclusion and Limitations

This paper presents a VDM approach that effectively integrates events and frames. Our method demonstrates high accuracy in measuring large-scale deformations and displacements in challenging scenarios, significantly advancing the capability and applicability of visual deformation measurement. Nevertheless, our method has certain limitations. Relying on **spatial continuity** and **brightness constancy** assumptions, it may introduce pseudo-strains when the target undergoes topological changes and suffer from performance degradation in scenarios with specular reflections, as shown in Fig. 8. In future work, we aim to address these limitations, improve system robustness, and explore 3D event-based VDM approaches[8, 21, 55].

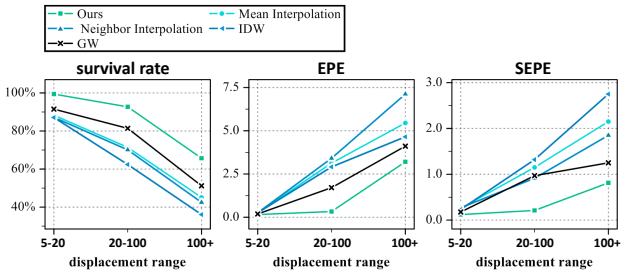

Figure 7. **Ablation study of displacement interpolation methods.** Performance comparison of our affine-invariant interpolation method against mean interpolation, neighborhood interpolation, Inverse Distance Weighting (IDW), and Gaussian Weighted Interpolation (GWI). Our method demonstrates superior measurement accuracy across different interpolation strategies.

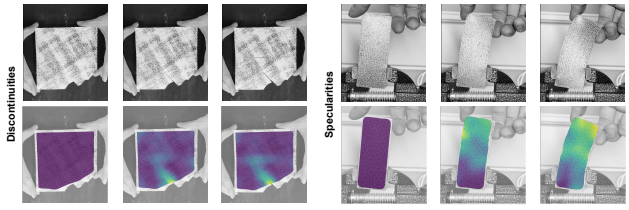

Figure 8. Failure cases of our system are demonstrated in scenarios involving plasterboard cracking and specular reflection.

Acknowledgments. This work is supported by the National Natural Science Foundation of China (NSFC) under Grants 62225207, 62436008, 62306295, and 62576328. The AI-driven experiments, simulations and model training were performed on the robotic AI-Scientist platform of Chinese Academy of Sciences.

References

[1] Antonio Agudo, Lourdes Agapito, Begona Calvo, and Jose MM Montiel. Good vibrations: A modal analysis approach for sequential non-rigid structure from motion. In *Proceedings of the IEEE Conference on computer vision and pattern recognition*, pages 1558–1565, 2014. 3

[2] Antonio Agudo, Francesc Moreno-Noguer, Begoña Calvo, and José María Martínez Montiel. Sequential non-rigid structure from motion using physical priors. *IEEE transactions on pattern analysis and machine intelligence*, 38(5): 979–994, 2015. 3

[3] Seyfeddine Boukhtache, Kamel Abdelouahab, François Berry, Benoît Blaysat, Michel Grediac, and Frédéric Sur. When deep learning meets digital image correlation. *Optics and Lasers in Engineering*, 136:106308, 2021. 6, 7

[4] S Boukhtache, K Abdelouahab, A Bahou, F Berry, B Blaysat, M Grédiac, and Frédéric Sur. A lightweight convolutional neural network as an alternative to dic to measure in-plane displacement fields. *Optics and lasers in engineering*, 161:107367, 2023. 3

[5] Christian Brandli, Raphael Berner, Minhao Yang, Shih-Chii Liu, and Tobi Delbruck. A 240× 180 130 db 3 μs latency global shutter spatiotemporal vision sensor. *IEEE Journal of Solid-State Circuits*, 49(10):2333–2341, 2014. 3

[6] Runyu Cao, Wen Xiao, Feng Pan, Ran Tian, Xintong Wu, and Lianwen Sun. Displacement and strain mapping for osteocytes under fluid shear stress using digital holographic microscopy and digital image correlation. *Biomedical Optics Express*, 12(4):1922–1933, 2021. 1, 2

[7] Yutian Chen, Shi Guo, Fangzheng Yu, Feng Zhang, Jinwei Gu, and Tianfan Xue. Event-based motion magnification. *arXiv preprint arXiv:2402.11957*, 2024. 3

[8] Hoonhee Cho, Jegyeong Cho, and Kuk-Jin Yoon. Learning adaptive dense event stereo from the image domain. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17797–17807, 2023. 8

[9] Guillermo Gallego, Henri Rebecq, and Davide Scaramuzza. A unifying contrast maximization framework for event cameras, with applications to motion, depth, and optical flow estimation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3867–3876, 2018. 3

[10] Guillermo Gallego, Tobi Delbrück, Garrick Orchard, Chiara Bartolozzi, Brian Taba, Andrea Censi, Stefan Leutenegger, Andrew J Davison, Jörg Conradt, Kostas Daniilidis, et al. Event-based vision: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 44(1):154–180, 2020. 3

[11] Daniel Gehrig and Davide Scaramuzza. Low-latency automotive vision with event cameras. *Nature*, 629(8014):1034–1040, 2024. 3

[12] Daniel Gehrig, Henri Rebecq, Guillermo Gallego, and Davide Scaramuzza. Eklt: Asynchronous photometric feature tracking using events and frames. *International Journal of Computer Vision*, 128(3):601–618, 2020. 3

[13] Mathias Gehrig, Mario Millhäusler, Daniel Gehrig, and Davide Scaramuzza. E-raft: Dense optical flow from event cameras. In *2021 International Conference on 3D Vision (3DV)*, pages 197–206. IEEE, 2021. 3, 6, 7

[14] Mathias Gehrig, Manasi Muglikar, and Davide Scaramuzza. Dense continuous-time optical flow from event cameras. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(7):4736–4746, 2024. 3

[15] Katia Genovese, YU Lee, AY Lee, and JD Humphrey. An improved panoramic digital image correlation method for vascular strain analysis and material characterization. *Journal of the mechanical behavior of biomedical materials*, 27: 132–142, 2013. 2

[16] Jesse Hagenaars, Federico Paredes-Vallés, and Guido De Croon. Self-supervised learning of event-based optical flow with spiking neural networks. *Advances in Neural Information Processing Systems*, 34:7167–7179, 2021. 4

[17] Friedhelm Hamann, Suman Ghosh, Ignacio Juarez Martinez, Tom Hart, Alex Kacelnik, and Guillermo Gallego. Low-power continuous remote behavioral localization with event cameras. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18612–18621, 2024. 3

[18] Friedhelm Hamann, Ziyun Wang, Ioannis Asmanis, Kenneth Chaney, Guillermo Gallego, and Kostas Daniilidis. Motion-prior contrast maximization for dense continuous-time motion estimation. In *European Conference on Computer Vision*, pages 18–37. Springer, 2024. 3, 4

[19] Han Han, Wei Zhai, Yang Cao, Bin Li, and Zheng-jun Zha. Event-based tracking any point with motion-augmented temporal consistency. *arXiv preprint arXiv:2412.01300*, 2024. 3

[20] Xindang He, Run Zhou, Zheyuan Liu, Suliang Yang, Ke Chen, and Lei Li. Review of research progress and development trend of digital image correlation. *Multidiscipline Modeling in Materials and Structures*, 20(1):81–114, 2024. 3

[21] Diego Hitzges, Suman Ghosh, and Guillermo Gallego. Derdnet: Learning depth from event-based ray densities. *arXiv preprint arXiv:2504.15863*, 2025. 8

[22] Jianyong Huang, Xiaochang Pan, Xiaoling Peng, Tao Zhu, Lei Qin, Chunyang Xiong, and Jing Fang. High-efficiency cell–substrate displacement acquisition via digital image correlation method using basis functions. *Optics and lasers in engineering*, 48(11):1058–1066, 2010. 2

[23] Zhenyu Jiang. Opencorr: An open source library for research and development of digital image correlation. *Optics and Lasers in Engineering*, 165:107566, 2023. 6, 7

[24] Navami Kairanda, Edith Tretschk, Mohamed Elgharib, Christian Theobalt, and Vladislav Golyanik. f-sft: Shape-from-template with a physics-based deformation model. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3948–3958, 2022. 3

[25] Nikita Karaev, Yuri Makarov, Jianyuan Wang, Natalia Neverova, Andrea Vedaldi, and Christian Rupprecht. Cotracker3: Simpler and better point tracking by pseudo-labelling real videos. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6013–6022, 2025. 6, 7

[26] Yekutiel Katz and Zohar Yosibash. New insights on the proximal femur biomechanics using digital image correlation. *Journal of Biomechanics*, 101:109599, 2020. 1, 2

[27] Joel David Krehbiel, John Lambros, JA Viator, and Nancy R Sottos. Digital image correlation for improved detection of basal cell carcinoma. *Experimental Mechanics*, 50:813–824, 2010. 2

[28] Bohao Liao, Wei Zhai, Zengyu Wan, Zhixin Cheng, Wenfei Yang, Tianzhu Zhang, Yang Cao, and Zheng-Jun Zha. Ef-3dgs: Event-aided free-trajectory 3d gaussian splatting. *arXiv preprint arXiv:2410.15392*, 2024. 3

[29] Chuankun Liu and Ya Wei. Experimental investigation on damage of concrete beam embedded with sensor using acoustic emission and digital image correlation. *Construction and Building Materials*, 423:135887, 2024. 1

[30] YX Luo and YL Dong. Strain measurement at up to 3000° c based on ultraviolet-digital image correlation. *NDT & E International*, 146:103155, 2024. 1

[31] Abed Malti and Cédric Herzet. Elastic shape-from-template with spatially sparse deforming forces. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3337–3345, 2017. 3

[32] S Mguil-Touchal, F Morestin, and M Brunei. Various experimental applications of digital image correlation method. *WIT Transactions on Modelling and Simulation*, 17, 2024. 3

[33] Christen Millerdurai, Diogo Luvizon, Viktor Rudnev, André Jonas, Jiayi Wang, Christian Theobalt, and Vladislav Golyanik. 3d pose estimation of two interacting hands from a monocular event camera. In *2024 International Conference on 3D Vision (3DV)*, pages 291–301. IEEE, 2024. 3

[34] Mohammad Mostafavi, Lin Wang, and Kuk-Jin Yoon. Learning to reconstruct hdr images from events, with applications to depth and flow prediction. *International Journal of Computer Vision*, 129(4):900–920, 2021. 3

[35] Jalees Nehvi, Vladislav Golyanik, Franziska Mueller, Hans-Peter Seidel, Mohamed Elgharib, and Christian Theobalt. Differentiable event stream simulator for non-rigid 3d tracking. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1302–1311, 2021. 3

[36] Marco Palanca, Gianluca Tozzi, and Luca Cristofolini. The use of digital image correlation in the biomechanical area: a review. *International biomechanics*, 3(1):1–21, 2016. 2

[37] Federico Paredes-Vallés, Kirk YW Scheper, Christophe De Wagter, and Guido CHE De Croon. Taming contrast maximization for learning sequential, low-latency, event-based optical flow. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 9695–9705, 2023. 4

[38] Shien Ri, Jiaxing Ye, Nobuyuki Toyama, and Norihiko Ogura. Drone-based displacement measurement of infrastructures utilizing phase information. *Nature Communications*, 15(1):395, 2024. 1

[39] Teresa Serrano-Gotarredona and Bernabé Linares-Barranco. A 128× 128 1.5% contrast sensitivity 0.9% fpn 3 μs latency 4 mw asynchronous frame-free dynamic vision sensor using transimpedance preamplifiers. *IEEE Journal of Solid-State Circuits*, 48(3):827–838, 2013. 3

[40] Shintaro Shiba, Yannick Klose, Yoshimitsu Aoki, and Guillermo Gallego. Secrets of event-based optical flow, depth and ego-motion estimation by contrast maximization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(12):7742–7759, 2024. 3

[41] Timo Stoffregen and Lindsay Kleeman. Event cameras, contrast maximization and reward functions: An analysis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12300–12308, 2019. 3

[42] Stepan Tulyakov, Daniel Gehrig, Stamatios Georgoulis, Julius Erbach, Mathias Gehrig, Yuanyou Li, and Davide Scaramuzza. Time lens: Event-based video frame interpolation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 16155–16164, 2021. 3, 6, 7

[43] Stepan Tulyakov, Alfredo Bochicchio, Daniel Gehrig, Stamatios Georgoulis, Yuanyou Li, and Davide Scaramuzza. Time lens++: Event-based frame interpolation with parametric non-linear flow and multi-scale fusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17755–17764, 2022. 3

[44] Stefaan W Verbruggen, Myles J Mc Garrigle, Matthew G Haugh, Muriel C Voisin, and Laoise M McNamara. Altered mechanical environment of bone cells in an animal model of short-and long-term osteoporosis. *Biophysical journal*, 108 (7):1587–1598, 2015. 2

[45] Zengyu Wan, Ganchao Tan, Yang Wang, Wei Zhai, Yang Cao, and Zheng-Jun Zha. Event-based optical flow via transforming into motion-dependent view. *IEEE Transactions on Image Processing*, 33:5327–5339, 2024. 3

[46] Zengyu Wan, Wei Zhai, Yang Cao, and Zhengjun Zha. Emotive: Event-guided trajectory modeling for 3d motion estimation. *arXiv preprint arXiv:2503.11371*, 2025. 3

[47] Yin Wang and Jiaqing Zhao. Dic-net: Upgrade the performance of traditional dic with hermite dataset and convolution neural network. *Optics and Lasers in Engineering*, 160: 107278, 2023. 3

[48] Yuliang Wu, Ganchao Tan, Jinze Chen, Wei Zhai, Yang Cao, and Zheng-Jun Zha. Event-based asynchronous hdr imaging by temporal incident light modulation. *Optics Express*, 32 (11):18527–18538, 2024. 3

[49] Yuxuan Xue, Haolong Li, Stefan Leutenegger, and Joerg Stueckler. Event-based non-rigid reconstruction from contours. *arXiv preprint arXiv:2210.06270*, 2022. 3

[50] Yuxuan Xue, Haolong Li, Stefan Leutenegger, and Jörg Stückler. Event-based non-rigid reconstruction of low-rank parametrized deformations from contours. *International Journal of Computer Vision*, 132(8):2943–2961, 2024. 3

[51] Jiashuai Yang, Kemao Qian, and Lianpo Wang. R3-dicnet: an end-to-end recursive residual refinement dic network for larger deformation measurement. *Optics Express*, 32(1): 907–921, 2023. 3

[52] Ru Yang, Yang Li, Danielle Zeng, and Ping Guo. Deep dic: Deep learning-based digital image correlation for end-to-end displacement and strain measurement. *Journal of Materials Processing Technology*, 302:117474, 2022. 3

[53] Yixin Yang, Jin Han, Jinxiu Liang, Imari Sato, and Boxin Shi. Learning event guided high dynamic range video reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13924–13934, 2023. 3

[54] Sungsik Yoon, Hyung-Jo Jung, JC Knowles, and Hae-Hyoung Lee. Digital image correlation in dental materials and related research: A review. *dental materials*, 37(5):758–771, 2021. 1, 2

[55] Haimei Zhao, Jing Zhang, Zhuo Chen, Bo Yuan, and Dacheng Tao. On robust cross-view consistency in self-supervised monocular depth estimation. *Machine Intelligence Research*, 21(3):495–513, 2024. 8

[56] Yang Zheng, Adam W Harley, Bokui Shen, Gordon Wetzstein, and Leonidas J Guibas. Pointodyssey: A large-scale synthetic dataset for long-term point tracking. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 19855–19865, 2023. 6

[57] Boran Zhou, Suraj Ravindran, Jahid Ferdous, Addis Kidane, Michael A Sutton, and Tarek Shazly. Using digital image correlation to characterize local strains on vascular tissue specimens. *Journal of visualized experiments: JoVE*, 107, 2016. 2

[58] Alex Zihao Zhu, Liangzhe Yuan, Kenneth Chaney, and Kostas Daniilidis. Ev-flownet: Self-supervised optical flow estimation for event-based cameras. *arXiv preprint arXiv:1802.06898*, 2018. 3

[59] Yunhao Zou, Ying Fu, Tsuyoshi Takatani, and Yinqiang Zheng. Eventhdr: From event to high-speed hdr videos and beyond. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024. 3