

Data-efficient optical 4D measurements of moving objects with stereo event cameras and tailored structured illumination

Christoph Freitag^{1,2}; Aliya Turelli Khusainova¹; Dr. Stefan Heist¹; Dr. Andreas Stark³; Dr. Peter Kühmstedt¹; Univ.-Prof. Dr. rer. nat. Gunther Notni^{1,4}; Jun. Prof. Dr. Christian Franke^{3,5,6}

1. Fraunhofer-Institut für Angewandte Optik und Feinmechanik IOF,
Albert-Einstein-Str. 7, 07745 Jena,

2. Friedrich-Schiller-Universität Jena, Institut für Angewandte Physik,
Albert-Einstein-Str. 15, 07745 Jena,

3. Friedrich-Schiller-Universität Jena, Institut für Angewandte Optik und Biophysik,
Max-Wien-Platz 1, 07743 Jena,

4. Technische Universität Ilmenau, Fachgebiet Qualitätssicherung und Industrielle Bildverarbeitung,
Gustav-Kirchhoff-Platz 2, 98693 Ilmenau,

5. Friedrich-Schiller-Universität Jena, Jena Center for Soft Matter,
Philosophenweg 7, 07743 Jena,

6. Friedrich-Schiller-Universität Jena, Abbe Center of Photonics,
Albert-Einstein-Str. 6, 07745 Jena

Contact: +49 3641 807-244, christoph.freitag@iof.fraunhofer.de

Abstract

We present an optical 4D (3D+time) measurement system for data-efficient reconstruction of shape and motion of rigid objects. In contrast to frame-based active stereo systems, which repeatedly resample surfaces at high rates, our approach exploits event cameras to break the tight dependence of scanning rate to object speed. The system combines a stereo pair of event cameras with a projector that generates a periodically rotating contrast edge to sparsely probe the measurement volume. In contrast to frame-based cameras, event cameras only report brightness changes and thus adapt their measurements to the dynamics of the scene. The moving contrast edge sparsely probes the measurement volume. Triangulation of temporally corresponding events in both cameras yields 3D surface points with individual timestamps. We specifically target sub-scan dynamics, i.e. significant object motion during a single slow scan of the contrast edge. Allowing significant sub-scan motion fundamentally breaks the assumptions of frame-based structured-light systems and introduces two key challenges: ambiguous temporal correspondences and motion-induced 3D rolling-shutter distortions. Our reconstruction framework addresses both effects. First, a geometric 4D model of the structured illumination is estimated and used to resolve ambiguity. Second, we estimate the object's motion and warp all 4D points to a common reference time, compensating the rolling-shutter effect. Experiments with a translating rigid object demonstrate that, despite low scan rates and significant sub-scan motion, the proposed method recovers the object's shape with sub-millimetre deviation from a static reference.

Keywords: Event cameras, Structured illumination, Event-based stereo, 4D shape measurement, Motion compensation

Introduction

Optical 3D metrology offers numerous techniques for measuring surfaces of technical objects across different sizes and materials [1,2]. A large class of systems uses active triangulation, for example fringe projection or GOBO projection [3]. In these systems a “scan” is a time interval in which the entire measurement volume is sampled, either sequentially or in parallel. For moving objects, most frame-based sensors have to keep motion within a scan negligible to maintain temporal

consistency of projected patterns and camera exposures. This leads to high scan rates, high-speed cameras and powerful illumination [3].

For rigid bodies, the shape is constant, and the primary interest is motion in space. Frame-based active stereo sensors yet acquire dense 3D snapshots at high rates, repeatedly resampling the same surface. From a metrology perspective, this is redundant and therefore data-inefficient when only rigid motion is present.

Event cameras offer a complementary paradigm by acquiring measurements only where and when intensity changes occur [4,5]. Each pixel independently monitors brightness and emits an event when a contrast threshold is exceeded. The output is an asynchronous stream of events (x, y, p, t) that is sparse in space and time and naturally concentrates on regions where changes occur. When combined with structured illumination, this enables a type of 4D sensor that sparsely samples the measurement volume but with high temporal precision.

In this contribution we extend our previous event-based stereo 3D method for static scenes [6] to 4D measurements of moving rigid objects. A rotating contrast edge generated by a projector probes the scene, and two event cameras observe the resulting brightness changes. We explicitly allow sub-scan dynamics, i.e. significant motion during a single slow scan, and compensate the resulting distortions in the reconstruction.

Related works

Event-based vision has been reviewed extensively by Gallego et al. [4], and recent sensor developments are summarized by [7]. For 3D reconstruction, many methods don't actively sample scenes with structured illumination and single or stereo event cameras. They rather focus on sparse 3D reconstruction and sparse SLAM. These approaches are not designed for dense, metrology-grade surface measurements.

To obtain dense depth information, several authors have combined event cameras with active illumination. Muglikar et al. introduced event-based structured light using a scanning laser point [8]. Fu et al. encoded both temporal and spatial information in an event-based structured-light system [9]. The named systems use a single event camera and, to the best of our knowledge none of them explicitly separate projector-induced events from events caused by moving object features.

Our previous work [6] used a stereo pair of event cameras and a moving contrast edge to reconstruct dense depth maps of static scenes. Raw events were binned into event frames and processed with a classical stereo correlation algorithm [10]. This implicitly assumed negligible motion during a scan and enforced a

one-to-one correspondence between (sub)pixels in both event cameras. It did not represent the ambiguity that occurs when objects move during pattern projection, and it worked with a fixed frame rate.

The present work departs from these approaches by keeping the full temporal resolution of events (no frame binning), explicitly representing one-to-many temporal correspondences, resolving them via a 4D model of the projector, and finally compensating motion-induced 3D rolling-shutter distortions. To the best of our knowledge, no existing approach explicitly models and resolves the ambiguity between illumination-induced and motion-induced events in a unified 4D reconstruction framework.

Methodology

Principle of event-based structured illumination

Our setup consists of a stereo rig of two event cameras observing a measurement volume illuminated by a projector. Inside the projector, a back-side-illuminated rotating mask wheel produces a sharp transition between bright and dark illumination. This projected contrast edge sweeps periodically through the measurement volume.

When the image of the contrast edge passes over an event camera's pixel, the brightness changes abruptly and the pixel emits a burst of events (see Fig. 1). Static regions without changes in illumination remain silent except for noise. Thus, each scan of the contrast edge samples the measurement volume in thin 3D slices over time. Using calibrated stereo geometry, temporally corresponding bursts in both cameras are triangulated to obtain 3D points. The matched stimuli's times provide the 4th coordinate.

An important consequence is the decoupling of scan duration and temporal resolution. The contrast edge can scan slowly across the field of view, while individual object points are sampled as quickly as the contrast edge's image passes from on pixel to its neighbour. We exploit this to allow significant object motion during a single scan (sub-scan dynamics) and correct the arising spatial distortion in post-processing.

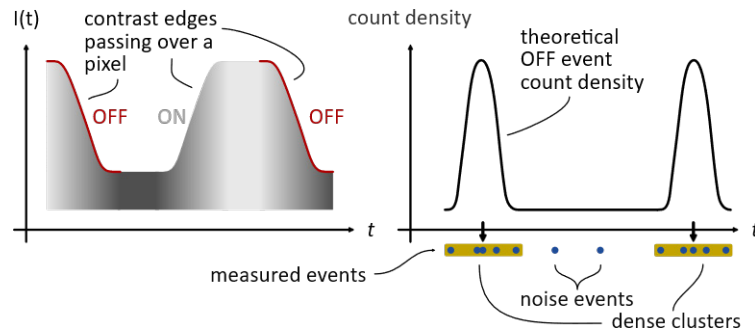


Fig. 1: How an event camera's pixel measures the passing of contrast edges. The left diagram shows the intensity above the pixel over time. Three contrast edges pass, of which two cause a decrease in brightness (OFF edges). Right: Measured events follow a density distribution. To distinguish signal events from noise events, they are clustered by density. Median times of clusters are defined as "stimuli" and pointed out by arrows on the timeline.

Event-generation model and detection of contrast-edge stimuli

The raw event streams contain bursts of events induced by the projector, events from moving object features, and electronic noise. To extract the temporal structure of the projected illumination without imposing an artificial frame rate, we detect contrast-edge stimuli for each pixel directly in the asynchronous event stream.

We model the event-generating process in a pixel by two categories. Burst-like stimuli appear when a contrast edge's image (from the projector or a moving object's feature) passes over a pixel. Between such bursts, there may be noise events. We use a clustering approach to find burst-like stimuli and separate from noise events. To achieve that, we compute the temporal differences between consecutive events. These differences form two classes: short intervals within bursts and longer intervals between bursts or noise events. A threshold duration that separates both classes can be found through different strategies. We use a two-class variance criterion based on Otsu's method [11]. Consecutive events separated by durations shorter than the threshold form a cluster, which is interpreted as a contrast-edge stimulus. The central (representative) stimulus time is defined as the median timestamp of the cluster. For each pixel, this yields a small set of stimulus times without defining frames or a fixed frame rate.

Stereo matching and 4D reconstruction

The next step is to find, for each stimulus in the primary camera, potential corresponding stimuli

in the secondary camera that might have originated from the same object point. We use calibrated intrinsic and extrinsic parameters (obtained with OpenCV [12]) to define epipolar curves in the original distorted image coordinates.

For each stimulus in a reference pixel of the primary camera: The corresponding epipolar curve in the secondary camera is computed. We slide a small three-pixel neighbourhood along the curve. In each of the three pixels we select the stimulus time closest to the reference time (if available). If all three pixels provide a stimulus, we fit a line $t(c) = v^{-1}c + t_0$ through the three (column c_i , time t_i) pairs. The slope v^{-1} represents the inverse image velocity of the contrast edge. If the temporal fit error is below a threshold and the back-projected column at the reference time lies within ± 0.5 pixel of the central column, we accept a match between the reference pixel and a sub-pixel coordinate in the secondary camera.

Because moving object edges can generate additional stimuli, a single stimulus in the primary camera can be matched to multiple candidate locations in the secondary camera. We keep this one-to-many ambiguity at this stage. Each accepted pair of corresponding pixels is undistorted and triangulated with the calibrated camera matrices to obtain a 3D point. The time coordinate of the resulting 4D point is the stimulus time in the primary camera. The raw 4D point cloud contains both true surface points and ghost points arising from false correspondences.

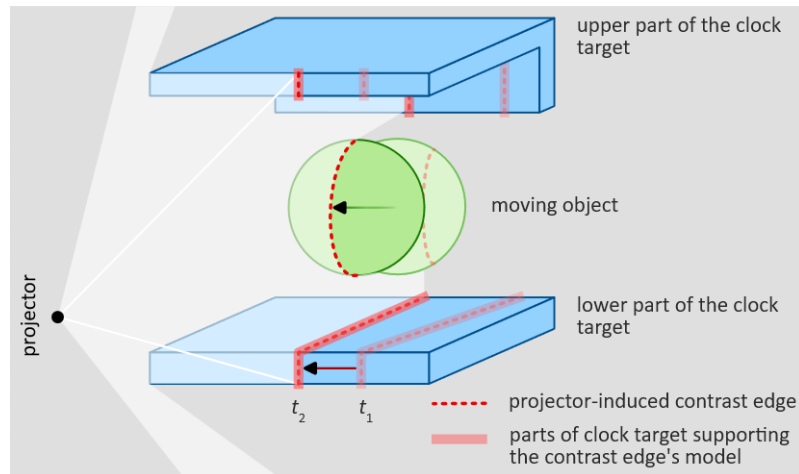


Fig. 2: Visualisation of the measurement volume to support the modelling of the structured illumination. The projector floods the measurement volume with two different levels of illumination, whose transition, a contrast plane, intersects with the objects' surfaces, creating a contrast edge (red). An object (green) is moving along the horizontal direction. Below and above the object, there is the static clock target (blue), whose unambiguous 4D reconstructed points are used to fit the model of the contrast plane which rotates around an axis at the projector's centre of projection. Two sources of events can be found in this scene: First, when the contrast edge's image passes over pixels of the event cameras, it triggers events. Second, images of the moving object's features with sufficient contrast (for instance, in this case the sphere's edge with different reflectivity than the background) trigger additional simultaneous events. Therefore, a horizontal temporal correspondence search on epipolar lines results in ambiguous matches.

Projector model and ghost-point filtering

To distinguish physically consistent 4D points from ambiguities, we introduce a geometric 4D model of the projector that enforces spatio-temporal consistency with the structured illumination (see Fig. 2). The plane rotates around an axis near the projector's centre of projection with approximately constant angular velocity. In the world coordinate system of the 4D points, the model has six parameters: two for the rotation axis' position on the xz plane, two for its orientation, one for angular speed, and one temporal offset.

A static "clock target" with known structure is placed in the measurement volume. During a scan, the rotating contrast plane intersects the clock target at several locations, providing 4D points that sample the plane. Since projector and cameras are not hardware-synchronised, we first detect scan periods using the temporal distribution of events on the clock target. Exploiting the periodic structure of the scans, the scan frequency and phase are estimated using a Rayleigh-type periodicity analysis [13].

For each scan (or scan segment), we then fit the projector model parameters to the 4D points belonging to the static clock target by

minimising their distances to the predicted contrast plane. The resulting model predicts which 3D locations are consistent with the projected contrast edge based on their time coordinates. All 4D points that deviate more than the fit's RMSD from the predicted plane at their timestamp are classified as ghost points and removed. This resolves most stereo ambiguities caused by moving object edges while also removing many noise 4D points.

Motion compensation

After ghost-point filtering, the remaining 4D points likely correspond to locations sampled by the projector's contrast edge. If the object moves during the scan, its surface points are sampled at different times, and the reconstructed shape exhibits a 3D rolling-shutter effect: it appears stretched along the motion direction.

We correct this distortion by explicitly modelling the temporal sampling process and estimating the object's motion to map all measurements to a common reference time t_0 . For a proof of concept, in our experiments we assume a single rigid object undergoing constant linear motion without rotation. Its displacement relative to its position at t_0 is described by

$$\Delta \mathbf{r}(t) = \mathbf{v}(t - t_0)$$

where $\mathbf{v}$ is the 3D velocity vector. Each 4D point $(\mathbf{r}_i, t_i)$ is mapped to

$$\mathbf{r}'_i = \mathbf{r}_i - \mathbf{v}(t_i - t_0)$$

and its time coordinate is set to t_0 . The resulting point cloud approximates the object shape at time t_0 and can be compared to a static reference.

To estimate $\mathbf{v}$ and t_0 , we first acquire a static reference point cloud with the object at rest. For a dynamic scan, we apply candidate motion parameters, warp the dynamic 4D points to t_0 , and compute, for each dynamic point, the distance to its nearest neighbour in the static point cloud. The root-mean-square of these distances serves as loss function. Since this loss function is not differentiable everywhere, we use a derivative-free optimisation method to minimise it and obtain $\mathbf{v}$ and t_0 . We took special care with non-overlapping parts between static and dynamic point clouds. Their associated large distances to points of the other point cloud can distort the loss function such that the optimisation produces erroneous estimates for $\mathbf{v}$ and t_0 .

Experimental setup

The experimental setup (Fig. 3 and Fig. 4) consists of a structured-light projector, a stereo rig of two event cameras, a static clock target, and a motorised linear stage in the measurement volume.

The projector is based on high-power LED illumination and a rotating mask wheel. The wheel comprises 80 masks. In the overlapping fields of view of the cameras only one mask image is visible, producing two contrast edges of opposite polarity. The cameras are configured to detect only one polarity (OFF events), effectively using a single contrast edge at a time. The contrast edge sweeps across the common field of view with an image velocity of approximately 2270 px s^{-1} , corresponding to about 265 mm s^{-1} at a mean object distance of 0.59 m . The scan frequency is about 1.6 Hz , so one scan takes roughly 600 ms .

The stereo rig holds two commercial event cameras with resolution of example 1280×720 pixels equipped with C-mount lenses of 25 mm focal length. The baseline and viewing geometry are chosen such that the projected contrast edge remains within the overlapping fields of view over the full scan. Camera biases

are tuned to ignore ON events and keep noise low in static regions.



Fig. 3: Measurement part of our experimental setup. Two event cameras observe a scene beyond the right edge. The mask wheel is illuminated from the light source within. Rotation causes the projected images of the mask to form moving contrast edges in the measurement volume. These sample the scene and trigger event measurements in the event cameras.

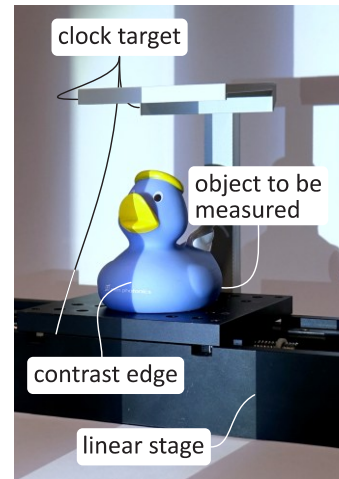


Fig. 4: A view into the measurement volume of our experimental setup with three elements: The duck, which is the object to be measured, a linear stage (black) and a clock target as reference for our contrast edge model.

A linear translation stage is used to move a rigid “duck” object horizontally through the measurement volume at controlled speeds. In the results shown below the object speed is approximately 10 mm s^{-1} , corresponding to about 85 px s^{-1} or a translation of about

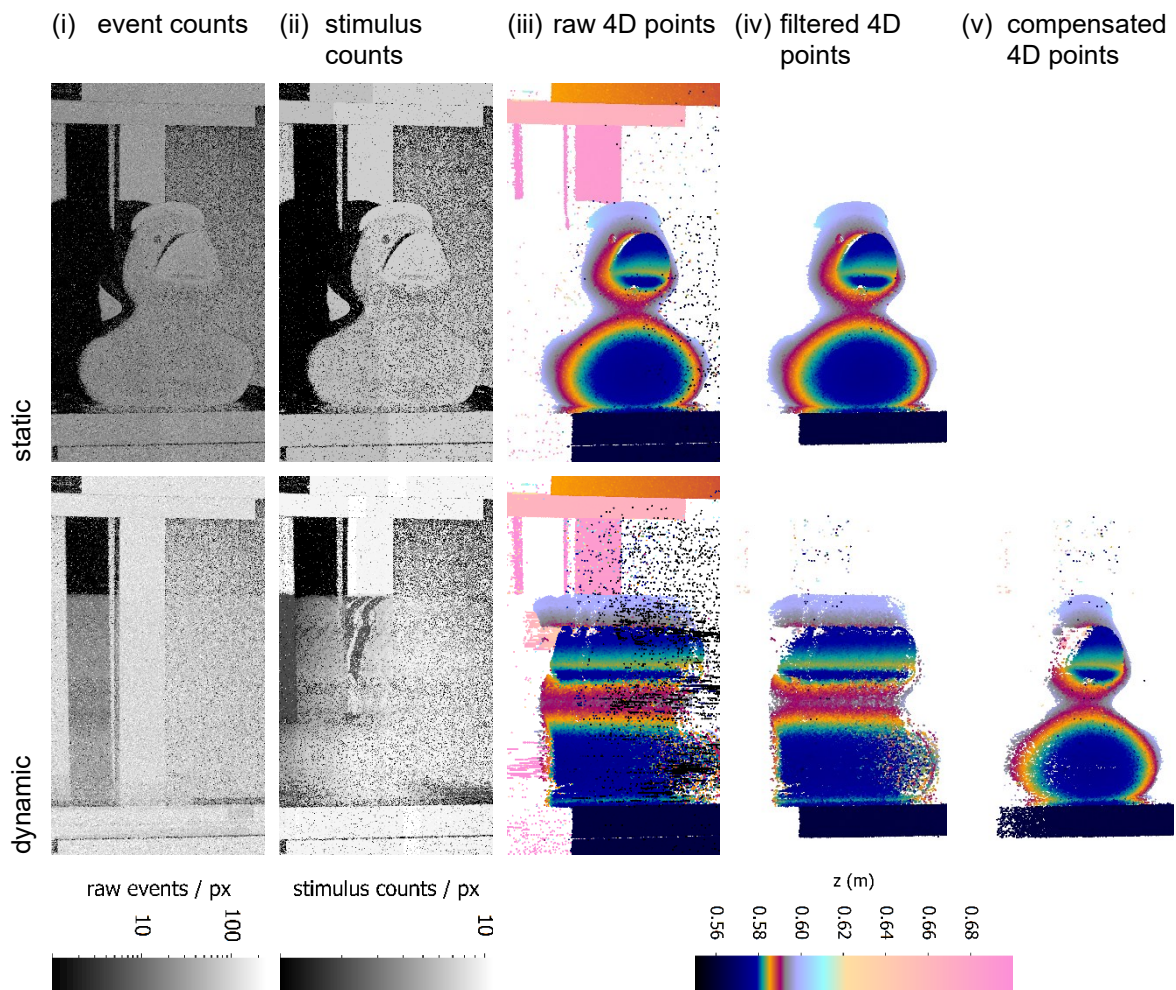


Fig. 5: Following the data from raw events (left) to the final motion-compensated point cloud (right). (i) The first column shows the total number (logarithmic) of events per pixel expressed in brightness. (ii) In the second row, raw events have been reduced to identified contrast edge stimuli per pixel. (iii) Ambiguously matched stimuli result in raw 4D points which consist of true object surface points and wrongly matched and triangulated ghost points. (iv) Using the model of the rotating contrast edge induced by the projector allows to remove ghost and noise points, resulting in a clean reference point cloud in the static case. The dynamic case shows multiple overlapping point clouds of the duck from multiple scans during motion. (v) The motion compensated point cloud is compared to the static reference.

53 pixels per scan in the camera's image plane. The static clock target spans the common field of view of both cameras. This is achieved by three horizontal planes, two at the top and one at the bottom of the measurement volume (see Fig. 2 and Fig. 4).

Results

Reconstruction pipeline

Fig. 5 illustrates the main processing stages for a static and a dynamic scene containing the duck object. For each case, the figure shows (i) event counts per pixel, (ii) the number of detected contrast-edge stimuli per pixel after clustering, (iii) raw 4D points after stereo triangulation, (iv) 4D points after projector-

based ghost-point filtering and removed clock target, and, (v) for the dynamic case, the motion-compensated 3D point cloud.

In the static case, projector-based filtering mainly removes a small number of wrongly matched 4D points. The shape of the duck is already largely correct in the raw 4D point cloud.

However, in the dynamic case, the raw 4D point cloud contains ghost points due to ambiguous temporal correspondences along the epipolar curves. These ghost points appear as spurious structures and violate the simple model of a rotating contrast plane. After fitting the projector model and rejecting inconsistent points, most ghost points are removed and a clean, but

motion-distorted 4D representation of the moving object remains.

Sub-scan dynamics and motion compensation

During a single scan of about 600 ms, the duck object moves roughly 6 mm, corresponding to more than 50 pixels in the image. This clearly violates the quasi-static assumption of classical active stereo systems. Nevertheless, the temporal resolution implied by the contrast-edge velocity on the sensor is on the order of 0.4 ms, many times finer than the scan duration.

If motion is not compensated, the filtered 4D point cloud of the dynamic scan exhibits a 3D rolling-shutter distortion: early points are measured when the object is on one side of the field of view, and later points when it has moved significantly, resulting in a stretched or contracted surface.

We acquired an additional static reference scan of the duck at rest. Using the procedure outlined in the Methodology section, we estimated the translational velocity vector and reference time t_0 that minimise the distance between the motion-compensated dynamic point cloud and the static reference. After compensation, the shape of the moving duck closely matches the static shape except for some sparsely recovered patches. Points are missed when Otsu's method does not yield an optimal separation threshold when finding stimuli in a pixel's raw events. The root-mean-square distance between static and compensated dynamic point clouds is about 0.3 mm when 1% of the largest residuals are discarded as outliers. Fig. 6 shows a spatial residual map over the duck's surface and the distribution of residuals. Larger residuals on parts of the translation platform are due to incomplete coverage of these areas in the static reference; these points were excluded from the optimisation.

From a metrology point of view, these results indicate that, despite low scan rates and substantial sub-scan motion, the event-based system yields shape deviations in the sub-millimetre range for the tested configuration.

Discussion

The presented method rethinks several stages of a structured-light 4D reconstruction pipeline for dynamic scenes based on actively probed event measurements. At the event level, a statistical model of the event-generating process motivates a clustering-based detection

of contrast-edge stimuli that preserves temporal precision without imposing a frame rate. Stereo matching is formulated as a one-to-many correspondence problem constrained by epipolar geometry and the local temporal behaviour of the moving contrast edge.

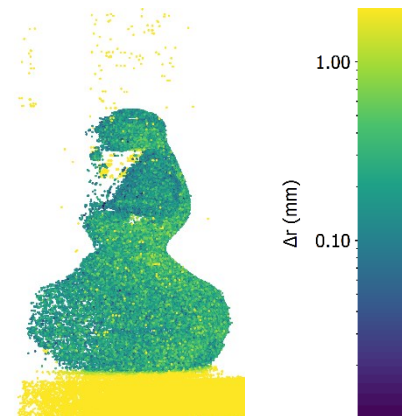


Fig. 6: 3D point cloud of the duck recovered from a dynamic measurement. Points are coloured as to reflect the distance to the nearest point of the static reference point cloud.

A key element is the 4D geometric model of the projector, which enables classification of 4D points as projector-consistent or inconsistent. This step removes ghost points originating from moving object edges and resolves ambiguities that cannot be eliminated in image space alone. Finally, a motion compensation stage treats the 3D rolling-shutter effect caused by sequential sampling of a moving surface.

In the current implementation, a static clock target is used as a calibration aid to identify contrast-edge events and fit the projector model. This component is not a fundamental requirement of the approach but a practical solution in the present proof-of-concept implementation. The clock target reduces the usable measurement volume and constrains the admissible motion directions. Moreover, the motion model is restricted here to a single rigid object undergoing constant linear motion, and an additional static scan is required for each object geometry.

Despite these limitations, the experimental results show that event-based structured light has the potential to make 4D measurements of moving rigid objects considerably more data-efficient than classical frame-based systems. We intend to test this with a thorough comparison in the future. Our sensor avoids high redundancy compared to frame-based 4D sensors when measuring rigid objects. This

makes the approach attractive for inline quality assurance and Industry 4.0 environments, where continuous 3D inspection of moving objects is required without interrupting the process.

Conclusions

We have presented a stereo event-based 4D measurement system with tailored structured illumination for data-efficient and physically consistent reconstruction of shape and motion of rigid objects. The system combines a slowly scanning contrast edge with the high temporal resolution of event cameras, detects contrast-edge stimuli by clustering raw events, resolves stereo ambiguities via a 4D projector model, and compensates the 3D rolling-shutter effect by estimating and inverting object motion.

References

- [1] T. Luhmann, S. Robson, S. Kyle, and J. Boehm, *Close-Range Photogrammetry and 3D Imaging* (De Gruyter, Berlin, Boston, 2020), <https://doi.org/10.1515/9783110607253>.
- [2] A. G. Marrugo, F. Gao, and S. Zhang, "State-of-the-art active optical techniques for three-dimensional surface metrology: a review," *J. Opt. Soc. Am. A* 37, B60–B77 (2020), <https://doi.org/10.1364/JOSAA.398644>.
- [3] S. Heist, P. Dietrich, M. Landmann, et al., "GOBO projection for 3D measurements at highest frame rates: a performance analysis," *Light: Science & Applications* 7, 71 (2018), <https://doi.org/10.1038/s41377-018-0072-3>.
- [4] G. Gallego, T. Delbrück, G. Orchard, et al., "Event-Based Vision: A Survey," *IEEE Trans. Pattern Anal. Mach. Intell.* 44, 154–180 (2022), <https://doi.org/10.1109/TPAMI.2020.3008413>.
- [5] T. Finateu, A. Niwa, D. Matolin, et al., "A 1280×720 Back-Illuminated Stacked Temporal Contrast Event-Based Vision Sensor with 4.86 μm Pixels, 1.066 GEPS Readout, Programmable Event-Rate Controller and Compressive Data-Formatting Pipeline," in *2020 IEEE International Solid-State Circuits Conference (ISSCC)*, pp. 112–114 (2020), <https://doi.org/10.1109/ISSCC19947.2020.9063149>.
- [6] C. Freitag, S. Heist, A. Stark, et al., "Photometrically optimized event-based stereo 3D measurements," *Opt. Express* 33, 2924–2939 (2025), <https://doi.org/10.1364/OE.541659>.
- [7] B. Chakravarthi, A. A. Verma, K. Daniilidis, et al., "Recent Event Camera Innovations: A Survey," in *Computer Vision – ECCV 2024 Workshops*, A. Del Bue, C. Canton, J. Pont-Tuset, and T. Tommasi, eds. (Springer Nature Switzerland, Cham, 2025), pp. 342–376, https://doi.org/10.1007/978-3-031-92460-6_21.
- [8] M. Muglikar, G. Gallego, and D. Scaramuzza, "ESL: Event-based Structured Light," in *2021 International Conference on 3D Vision (3DV)*, pp. 1165–1174 (2021), <https://doi.org/10.1109/3DV53792.2021.00124>.
- [9] J. Fu, Y. Zhang, Y. Li, et al., "Fast 3D reconstruction via event-based structured light with spatio-temporal coding," *Opt. Express* 31, 44588–44602 (2023), <https://doi.org/10.1364/OE.507688>.
- [10] P. Dietrich, S. Heist, M. Landmann, et al., "BICOS—An Algorithm for Fast Real-Time Correspondence Search for Statistical Pattern Projection-Based Active Stereo Sensors," *Appl. Sci.* 9 (2019), <https://doi.org/10.3390/app9163330>.
- [11] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst. Man Cybern.* 9, 62–66 (1979), <https://doi.org/10.1109/TSMC.1979.4310076>.
- [12] G. Bradski, "The OpenCV Library," *Dr. Dobb's Journal of Software Tools* (2000).
- [13] T. Bai, "Methods of Periodicity Analysis: Relationship between the Rayleigh Analysis and a Maximum Likelihood Method," *The Astrophys. J.* 397, 584–590 (1992), <https://doi.org/10.1086/171816>.

On a translating test object, the method achieves sub-millimetre agreement between static and dynamic reconstructions while operating at a low scan rate and allowing significant motion within a scan. Future work will address more general motion models, reduce reliance on auxiliary calibration structures such as the clock target, and provide a quantitative comparison to high-speed frame-based systems.

Funding

This work was supported in part by the German Federal Ministry of Education and Research (BMBF, project 13N16890) and by the Thüringer Aufbaubank (project FGR 0060 KI-supER).