Supplemental Material: Shadow-Free Projection with Blur Mitigation on Dynamic, Deformable Surfaces

1 Multi-Projector-Multi-Camera Calibration

To enable precise spatial alignment between projectors and depth sensors, we implemented an automatic calibration procedure that establishes correspondences between the 2D image coordinates of multiple projectors and the 3D coordinates obtained from multiple depth sensors. We project an 8×6 checkerboard pattern onto a wooden board at 32 different positions: 16 positions at one height level and an additional 16 positions at a second height level (we used $\delta h = 20\,\text{cm}$). This dual-height configuration eliminates the ambiguity between field-of-view and distance parameters inherent in projector calibration.

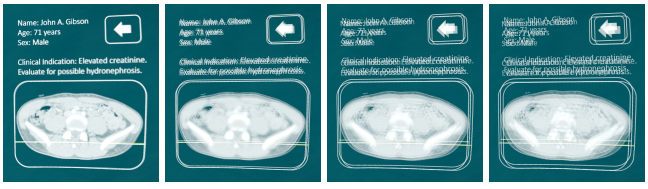
For each reference point i, we define the 2D image coordinates on projector j as $\mathbf{p}_i^{(j)} = [u_i^{(j)}, v_i^{(j)}]^T$, and measure the corresponding 3D point in the coordinate system of depth sensor k as $\mathbf{P}_i^{(k)} = [X_i^{(k)}, Y_i^{(k)}, Z_i^{(k)}]^T$. The calibration seeks to estimate the intrinsic matrix $\mathbf{K}^{(j)}$ of each projector, along with the extrinsic parameters $(\mathbf{R}^{(j,k)}, \mathbf{t}^{(j,k)})$ describing the rigid transformation between the sensor and projector coordinate systems. We model the projection using a standard pinhole formulation:

$$s_i \cdot \tilde{\mathbf{p}}_i^{(j)} = \mathbf{K}^{(j)} \cdot [\mathbf{R}^{(j,k)} | \mathbf{t}^{(j,k)}] \cdot \tilde{\mathbf{P}}_i^{(k)} \tag{1}$$

where $\tilde{\mathbf{p}}_i^{(j)} \in \mathbb{R}^3$ and $\tilde{\mathbf{P}}_i^{(k)} \in \mathbb{R}^4$ denote the homogeneous coordinates of the 2D and 3D points, respectively, and s_i is a scale factor. To align multiple depth sensors into a common coordinate system, corresponding 3D points $\mathbf{P}_i^{(\text{source})}$ and $\mathbf{P}_i^{(\text{target})}$ are registered using SVD-based rigid transformation estimation [1]. Reference points are automatically detected using the checkerboard detection algorithm by *Duda et al.* [2], and the system parameters are optimized via reprojection error minimization across all captured calibration poses. This approach enables largely automatic and precise calibration between multiple projectors and multiple depth cameras within minutes.

Implementation Note: For Ensenso N36 structured light sensors, the projected checkerboard patterns could be detected directly in the infrared image, while the Orbbec Femto required detection via the color image due to aggressive filtering of visible light in its IR camera.

2 Simulation of depth error



(a) 0 mm **(b) 1 mm** **(c) 2 mm** **(d) 3 mm**

Figure 1: Effect of depth sensing errors on a geometrically corrected UI projected by three projectors using a simulated constant depth offset. Results from our simulation software show that even small errors (1–3 mm) cause visible blurring.

To demonstrate the projection accuracy needed to prevent overlap blurring, we simulated a constant depth offset using our framework. The results are shown in Figure 1.

3 ToF-Multi-Path Interference Effects

(a) Default **(b) Scenario A** **(c) Scenario B** **(d) Scenario C**

Figure 2: Scenarios used to assess the impact of environmental interference in ToF sensors on projection precision.

Table 1: POD using ToF sensors for surface reconstruction: Average deviation of projected points between individual projectors, measured in millimeters. This metric reflects the perceived blurring caused by overlapping projections (without blur mitigation), as viewed from above.

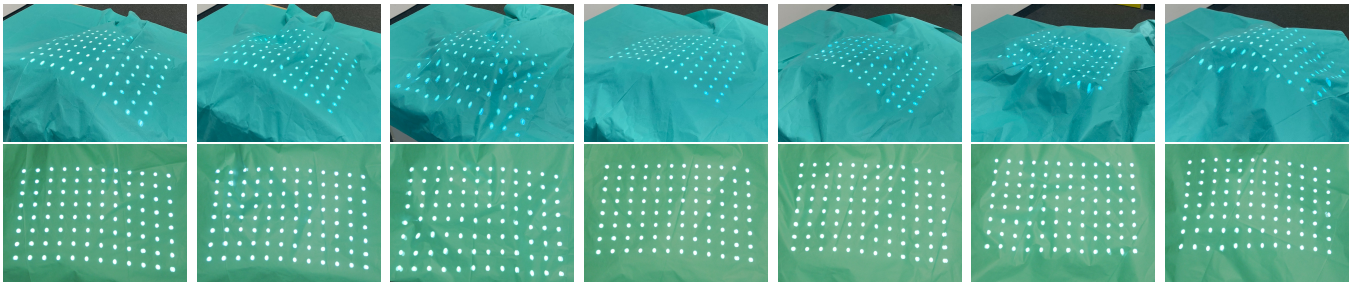
Surface	Projection Overlap Divergence (POD)			
Interference	Default	Scenario A	Scenario B	Scenario C
(1) Flat board	1.91	4.11	2.74	4.87
(2) Flat cloth	2.86	5.07	4.13	5.78
(3) Uneven cloth	15.28	19.38	16.95	20.61

Table 2: Deviation of identical projectors in different environmental scenarios—where people are positioned around the object—to the default scenario, measured in millimeters. ToF sensors are used for surface reconstruction.

Surface	Environment-dependent Interference Deviation to Default								
	Scenario A			Scenario B			Scenario C		
Proj.	1	2	3	1	2	3	1	2	3
(1)	1.62	2.10	3.06	0.73	1.08	1.62	2.28	2.51	3.61
(2)	1.35	1.78	2.68	0.57	0.90	1.26	2.24	2.15	3.48
(3)	1.77	2.27	3.66	0.73	0.99	1.58	2.43	2.80	4.31

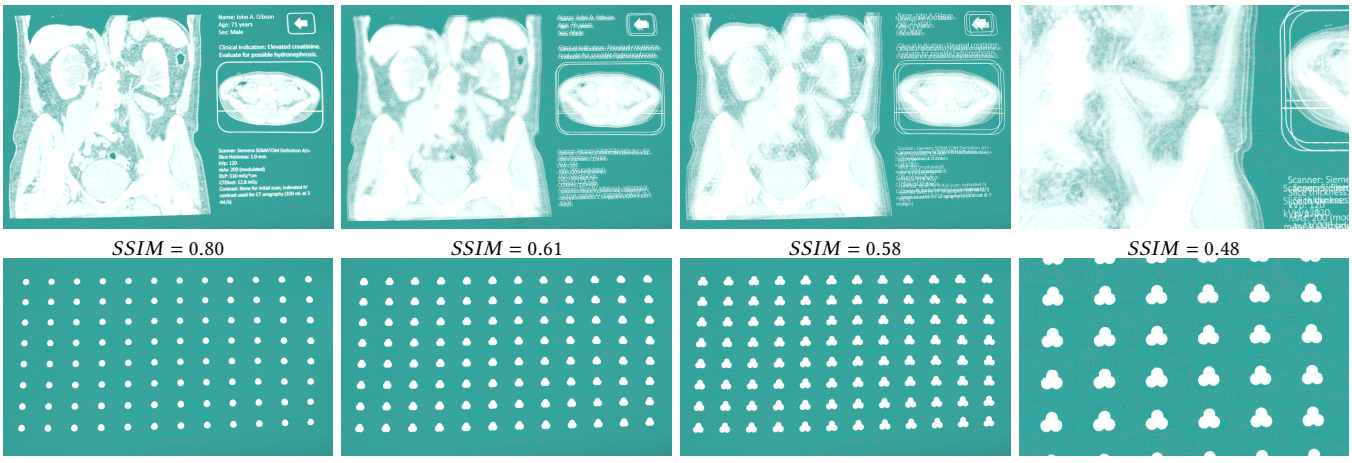
Time-of-Flight (ToF) sensors are known to suffer from depth inaccuracies due to environmental interference, material properties, and surface color [3–5]. We evaluated how these factors affect projector alignment precision using an Orbbec Femto Bolt sensor (comparable to Microsoft Azure Kinect).

Three surface types were tested: (a) flat wooden board, (b) flat surgical drape, and (c) wrinkled surgical drape. Each was evaluated



POD = 1.61 mm | POD = 1.60 mm | POD = 2.03 mm | POD = 1.35 mm | POD = 1.47 mm | POD = 1.49 mm | POD = 1.51 mm

Figure 3: Images of the uneven surfaces (N=7) used for precision measurement using two Ensenso N36 (SL) cameras (see paper, Table 1). Top: side view of the surface illustrating the uneven geometry. Middle: top-down view from the camera used for measurement. Bottom: separate measurements of the Projection Overlap Divergence $POD(P1, P2, P3)$. The reported error includes inaccuracies of calibration, depth sensor, and our surface reconstruction pipeline. For practical reasons, the measurement camera perspective did not exactly match the viewer's position, resulting in slight warping of the dot pattern in the middle row. As the measurements rely on inter-projector point distances, this does not significantly affect the results.



$SSIM$ = 0.80 | $SSIM$ = 0.61 | $SSIM$ = 0.58 | $SSIM$ = 0.48

POD = 0.5 mm | $SSIM$ = 0.96 | POD = 2.2 mm | $SSIM$ = 0.90 | POD = 3.9 mm | $SSIM$ = 0.89 | POD = 3.9 mm | $SSIM$ = 0.87

(a) Sim. depth error: 0 mm | **(b) Sim. depth error: 2 mm** | **(c) Sim. depth error: 4 mm** | **(d) Sim. depth error: 4 mm**

Figure 4: Comparison of SSIM and POD for evaluating the precision of overlapping projections in our simulation software. Depth errors of (a) 0 mm, (b) 2 mm, and (c–d) 4 mm were simulated, assuming an ideal calibration and ideal depth sensors. The virtual camera for capturing the images was positioned above the projection surface, consistent with the real-world comparison described in the paper. Two contents (UI and dot pattern) were projected onto a slightly uneven surface. Only images with all three overlapping projectors are shown here, but the metrics were computed pairwise on the individually projected images of each projector. Results: SSIM strongly depends on projected content and camera distance, and does not scale linearly with error. In contrast, POD is independent of projected content and camera distance and is related to the simulated error.

under four environmental conditions: without nearby people and with one to two people positioned around the setup (see Figure 2).

Results show that alignment error depends strongly on both surface and surroundings. The baseline average error (flat wood, no people) was 1.91 mm. Environmental interference increased errors up to 4.87 mm. Wrinkled surfaces introduced severe misalignments, with errors ranging from 15.28 mm to 20.61 mm.

Table 1 summarizes the *Projection Overlap Divergence (POD)*. Relative deviations across conditions for separate projectors are given in Table 2.

3.1 Further Precision Measurement Results

To provide a clearer understanding of projection accuracy on uneven surfaces, Figure 3 illustrates the results of seven trials conducted on uneven surfaces using the Ensenso N36 (SL) camera (see Table 1 in the paper), with their corresponding POD values. The results demonstrate that accurate projection can be achieved even on highly uneven surfaces.

To further highlight the difference between using SSIM and our proposed POD definition, Figure 4 presents both metrics on different examples. For controlled error evaluation, we used our simulation

environment to introduce artificial depth errors of 2 mm and 4 mm into the depth maps of simulated depth sensors. For SSIM, we used the displayed simulated camera images as input, while we detected the dot patterns on the simulated camera images using OpenCV for calculating the POD. Note that 96 dots were used in (Figure 4a–c), while only 24 dots were used in (Figure 4d). The results show that SSIM strongly depends on the projected content and the camera distance used for capturing the comparison images, and it does not scale linearly with the actual error. By contrast, our POD metric reflects these changes more reliably and is largely invariant to such external factors.

As seen in our setup (see teaser image, Figure 1 in the paper), projectors and depth cameras are tilted at roughly 40° relative to the projection surface, causing images to be cast at an angle. In such a configuration, a depth error (offset along the z-axis) is expected to result in a visible image shift by constant factor. This behavior is well captured by the POD metric, apart from a small offset: a 0 mm depth error corresponds to a POD of 0.5 mm, 2 mm to 2.2 mm, and 4 mm to 3.9 mm. Since the simulation assumes ideal calibration (with projector–camera transformations directly applied) and ideal depth sensors, the residual 0.5 mm error might be attributed to virtual sensor resolution limits (640 × 576) and discretization inaccuracies within the software pipeline.

4 Detailed Runtime Breakdown

To evaluate the runtime performance, we measured the timings of individual modules of our pipeline using a setup with three projectors (see Figure 5), each at a resolution of 3840×2160 while the depth sensors had an resolution of 640×576. In particular, we evaluated the effect of the number of depth sensors on the overall pipeline performance.

The OpenGL thread of our application — which includes geometric correction, shadow avoidance, GUI rendering via ImGUI, and debug visualization of the colored point cloud using BlendPCR in a 3D view — requires an average of 22.2 ms when using three depth cameras, 16.9 ms with two cameras, and 8.7 ms with a single camera. The debug visualization may be disabled to save roughly 1–2 ms of runtime.

During time measurement, up to three background threads were active, responsible for receiving live depth frames from the Orbbec sensors at a constant rate of 30 fps. Although our pipeline still offers potential for further optimization, the results clearly demonstrate that our blur mitigation approach enables high-performance, shadow- and blur-free projection.

Furthermore, we conducted a more detailed analysis of the runtime of individual components of the blur mitigation pipeline, see Figure 6. A more time-consuming part is the generation of the *Projector Distance Maps*, taking 1.18 ms for three depth cameras. These maps are implemented via adapted screen-space passes from BlendPCR and continuously render point clouds at a resolution of 1280×720.

Using the Jump-Flooding algorithm, we compute the *Shadow Distance Maps* for all three cameras in approximately 0.55 ms at a resolution of 256×256 pixels, which provides sufficient accuracy for shadow avoidance. The most computationally expensive step within the blur mitigation pipeline is the *Per-Tile Processing*, which takes

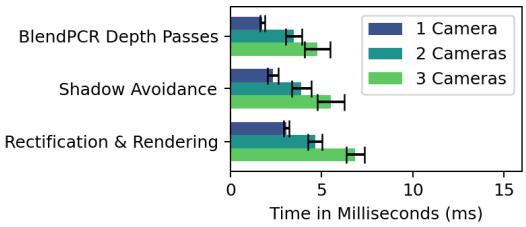


Figure 5: Runtime of modules in our geometric correction and shadow-free projection pipeline using 3 projectors at 3840×2160, depending on the number of depth cameras used.

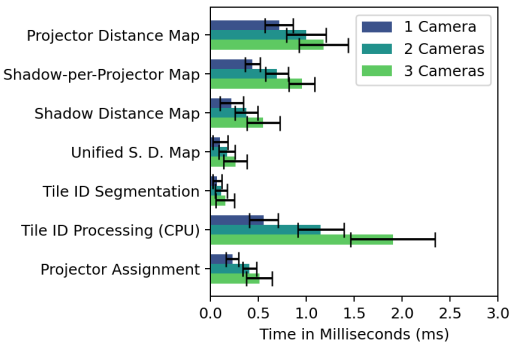


Figure 6: Runtime of individual components of the blur mitigation pipeline using 3 projectors, depending on the number of depth cameras used.

1.90 ms for three cameras. This step partly runs on the CPU and processes per-camera distance maps of size 256×256 to determine the minimum distances per tile.

5 Supplemental Materials for the User Study

Table 3: Effect sizes (Kendall's W and |r|) for the individual items of the NASA-TLX questionnaire, as well as for our three additional questions. Color shading indicates conventional thresholds for effect sizes: 0.1–0.3 (light green, small effect), 0.3–0.5 (orange, medium effect), and 0.5–1.0 (red, large effect).

	Kendall's W	Pairwise Wilcoxon (|r|)		
	No, A, B	No vs. A	No vs. B	A vs. B
Mental Demand	0.21	0.55	0.51	0.13
Physical Demand	0.02	0.41	0.32	0.21
Temporal Demand	0.01	0.01	0.14	0.04
Performance	0.21	0.51	0.52	0.22
Effort	0.19	0.6	0.25	0.26
Frustration	0.1	0.56	0.45	0.08
Char Readability	0.56	0.84	0.86	0.34
Plot Readability	0.48	0.88	0.6	0.44
Artifacts	0.38	0.85	0.49	0.28
NASA-TLX (Raw)	0.2	0.55	0.39	0.22

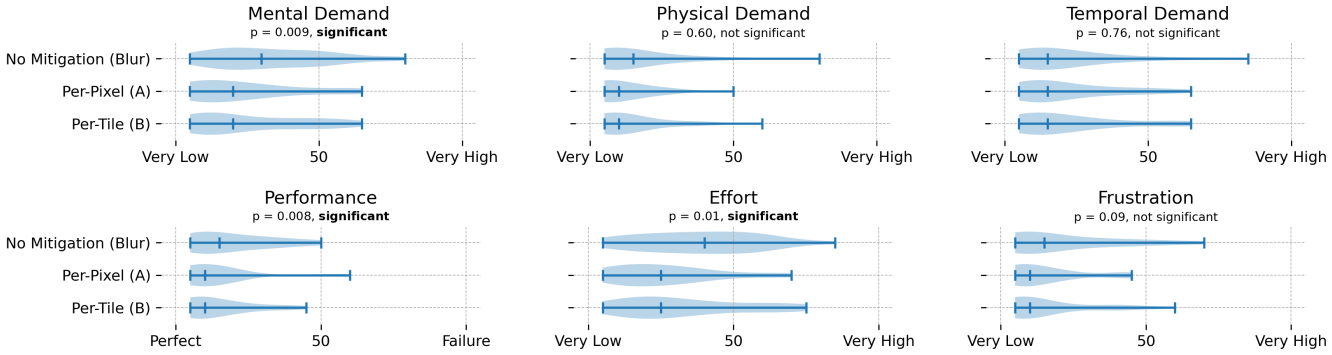


Figure 7: Detailed comparison of individual questions from the NASA-TLX test. A direct comparison of the NASA-TLX test questions reveals significant differences in *Mental Demand*, *Performance*, and *Effort*.

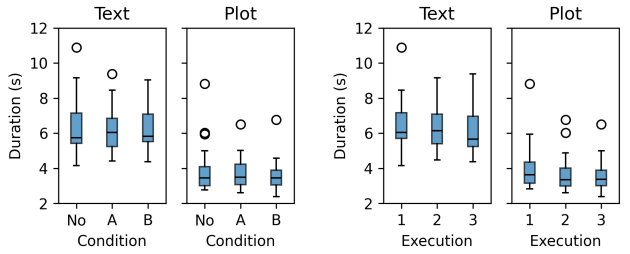


Figure 8: Results of the time measurements, cumulated as median values per participant. Left: grouped by condition; right: sorted by execution order. No significant differences are observed between conditions for texts ($p = 0.738$) and plots ($p = 0.84$). A minimal learning effect is evident, which is significant for plots ($p = 0.001$) but not for texts ($p = 0.119$).

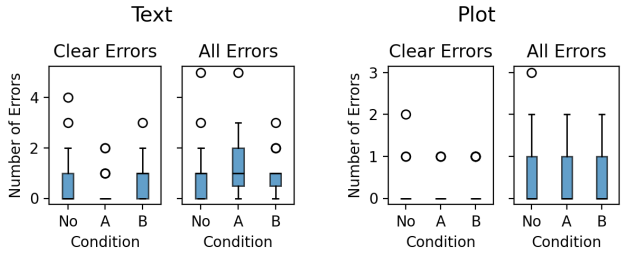


Figure 9: Results of the error measurements, reported as median values per participant. *Clear Errors* refer to mistakes that were unambiguously incorrect and not self-corrected by the participant, while *All Errors* include both clear errors and those that the participant immediately corrected. A slight trend is observable in the text data, suggesting that in *Condition Per-Pixel (A)*, fewer clear errors were made, but a higher number of self-corrected errors occurred. However, these results are not statistically significant. *P-values from left to right:* $p = [0.166, 0.664, 0.670, 0.790]$.

NASA-TLX questions and effect sizes. For completeness, we also report the results of the individual NASA-TLX questions (see Figure 7) as well as the effect sizes of all questions and the overall NASA-TLX score in Table 3.

Timings and Errors. In the user study, we also measured task completion times and observed errors; however, we did not find any statistically significant differences between the conditions, see Figure 8 and Figure 9. Still, there was a slight trend suggesting that participants made fewer clear reading errors in the *Per-Pixel Assignment (A)* condition, although this was accompanied by a slightly higher number of verbal slips or self-corrected errors.

The observed trends in the data — such as slightly fewer reading errors but more hesitations in *Per-Pixel Assignment (A)* — may be related to the faint, moving stripe artifact present in this condition. This artifact could have drawn participants' visual attention, potentially impacting reading fluency. These effects warrant further investigation to better understand the cognitive and perceptual implications of such subtle visual disturbances.

Reference to Main Paper

Andre Muehlenbrock, Yaroslav Purgin, Nicole Steinke, Verena Uslar, Dirk Weyhe, Rene Weller, and Gabriel Zachmann. 2025. Shadow-Free Projection with Blur Mitigation on Dynamic, Deformable Surfaces. In *31st ACM Symposium on Virtual Reality Software and Technology (VRST '25), November 12–14, 2025, Montreal, QC, Canada.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3756884.3766018

References

[1] K. S. Arun, T. S. Huang, and S. D. Blostein. 1987. Least-Squares Fitting of Two 3-D Point Sets. *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-9, 5 (1987), 698–700. doi:10.1109/TPAMI.1987.4767965

[2] Alexander Duda and Udo Frese. 2018. Accurate Detection and Localization of Checkerboard Corners for Calibration. In *Proc. of the British Machine Vision Conference.*

[3] S.A. Guomundsson, Henrik Aanæs, and Rasmus Larsen. 2007. Environmental Effects on Measurement Uncertainties of Time-of-Flight Cameras. 1 – 4. doi:10.1109/ISSCS.2007.4292664

[4] Andre Mühlenbrock, Roland Fischer, Christoph Schröder-Dering, René Weller, and Gabriel Zachmann. 2022. Fast, accurate and robust registration of multiple depth sensors without need for RGB and IR images. *The Visual Computer* (17 May 2022). doi:10.1007/s00371-022-02505-2

[5] Michal Tölgyessy, Martin Dekan, Ľuboš Chovanec, and Peter Hubinský. 2021. Evaluation of the Azure Kinect and Its Comparison to Kinect V1 and Kinect V2. *Sensors* 21, 2 (2021). doi:10.3390/s21020413