Fraunhofer
Heinrich Hertz Institute

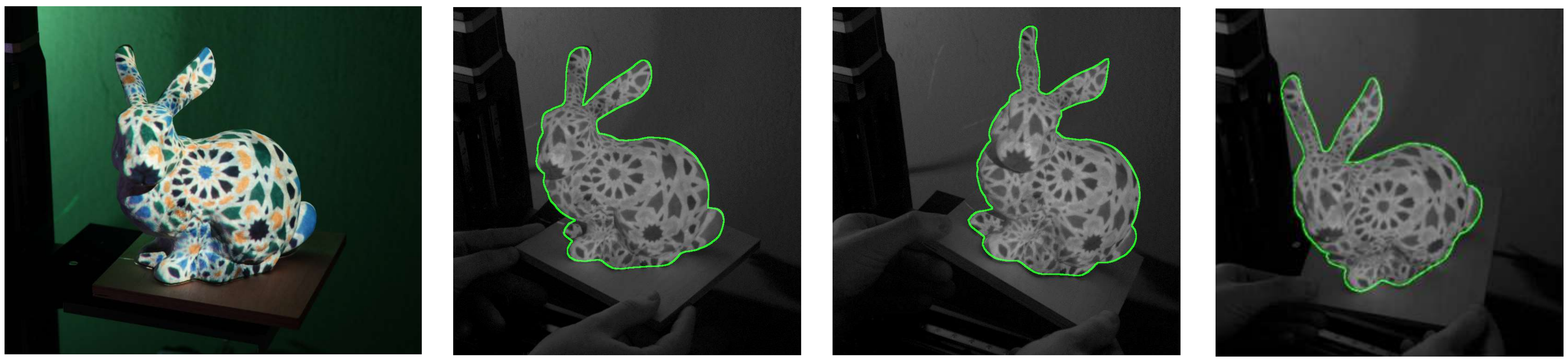
Computer Vision & Graphics
Vision & Imaging Technologies Department

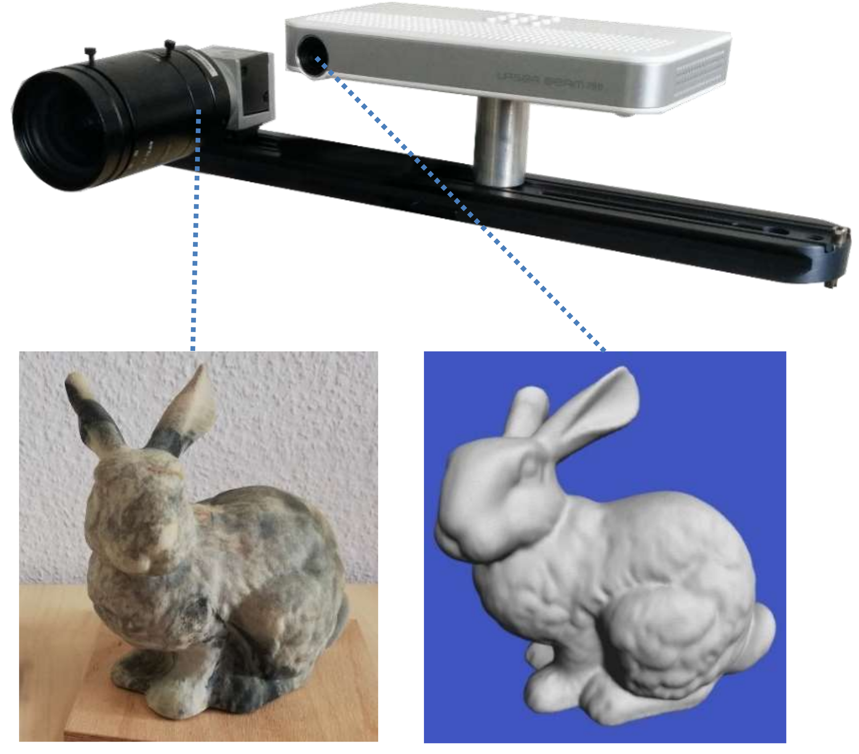
Visual Computing
Department of Computer Science

Humboldt-Universität zu Berlin

Projection Distortion-based Object Tracking in Shader Lamp Scenarios

Niklas Gard, Anna Hilsmann, Peter Eisert





(1) Problem Definition

- Projector-camera systems augment environment by projecting new textures on known target geometries.
- Shader lamps → stereo system: projector as inverted camera.
- Moving object *under* projection destroys illusion and distracts image-based trackers using contour or texture information.
- **Target:** Use local texture distortion of projection on surface in camera image for pose offset determination.

(2) Projection Distortion-based Pose Estimation

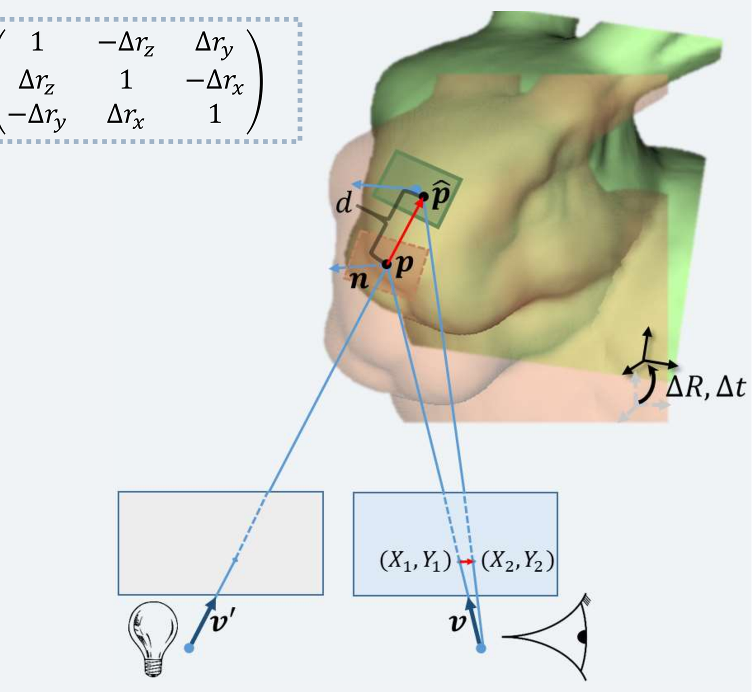
- Estimate optical-flow between camera image I and synthetically rendered image $\hat{I}$ → **analysis-by-synthesis** approach [1].
- $\boldsymbol{p}$ moves along viewing ray $\boldsymbol{v}'$ of projector to intersection with local plane moved with offset $\Delta\boldsymbol{R}$, $\Delta\boldsymbol{t}$.

$$\hat{\boldsymbol{p}} = \boldsymbol{p} + \boldsymbol{v}'d \quad d = \frac{(\Delta\boldsymbol{R}(\boldsymbol{p}-\boldsymbol{t}) + \Delta\boldsymbol{t} - (\boldsymbol{p}-\boldsymbol{t}))^T \Delta\boldsymbol{R}\boldsymbol{n}}{\boldsymbol{v}'^T \Delta\boldsymbol{R}\boldsymbol{n}} \quad \Delta\boldsymbol{R} = \begin{pmatrix} 1 & -\Delta r_z & \Delta r_y \\ \Delta r_z & 1 & -\Delta r_x \\ -\Delta r_y & \Delta r_x & 1 \end{pmatrix}$$

- Pixel-based equation system is solved for pose offset. Linearization of movement in image space.

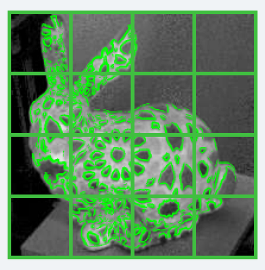
$$\frac{\partial \hat{I}}{\partial X} u_m + \frac{\partial \hat{I}}{\partial Y} v_m \approx a \begin{pmatrix} \Delta r \\ \Delta t \end{pmatrix} \approx \hat{I} - I$$

- $\boldsymbol{a} = (a_1, a_2, a_3, a_4, a_5, a_6)$ depends on
  - camera intrinsics and extrinsics,
  - normal map and depth map,
  - image derivative in x and y direction.



How to adapt ΔR,Δt to synthesize observation?

- Bridge domain gap with **edge images** and illumination simulation.
- **Stabilization** with Iterative Reweighted Least Squares.
- Image pyramid to allow larger pose offsets.



Adaptive local thresholds for edge detection.

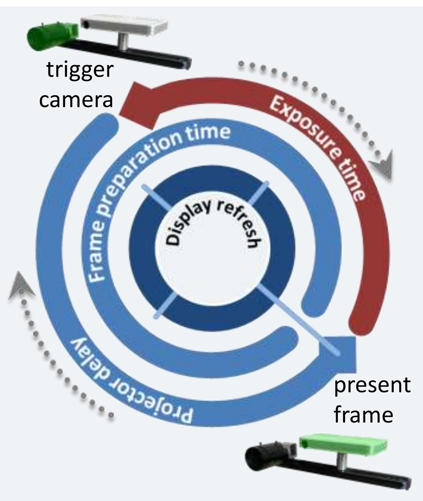
Initialization

- Pose initialization with silhouette-based tracker.
- Initial approximation of ambient illumination (Lambert, 5 DoF) and projector brightness (linear attenuation, 1 DoF).

Synthetic rendering vs. camera image for ambient illumination, white and texture projection (from left to right).

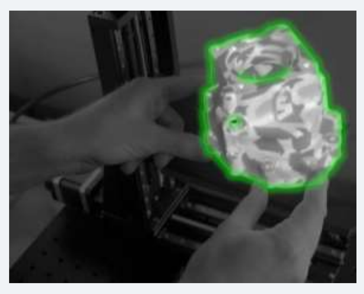
(3) Synchronization and Timing

- Synchronize exposure time to color wheel rotation.
- Parallelize on GPU for short frame preparation time.
- Optional: Synchronize projection with V-Blank for fastest performance.

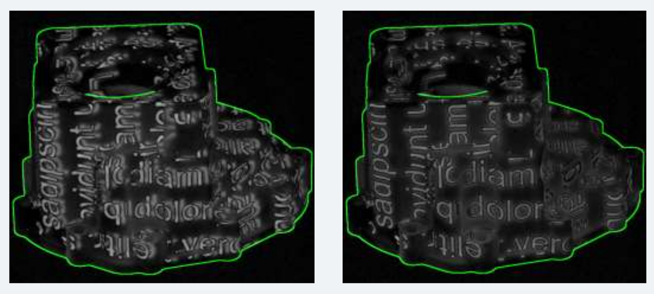


(4) Extensions

1. **Combined approach:**
   Use silhouette-based terms [2] near object border to prevent ambiguous configurations.
2. **Fast refinement of calibration:**
   Modify model to optimize extrinsic pose between projector and camera given the object pose in camera frame.



Silhouette-based terms are used near contour.



Improved matching between synthetic overlay and observation after calibration refinement.

(5) Results

Synthetic Data

- **Test 1**: continuous movement, small offsets per frame.
- **Test 2**: large abrupt offsets, multiple frames compensation.

		Text projection*			Colored pattern projection*		
t_{offset}	r_{offset}	t_{err}	r_{err}	valid**	t_{err}	r_{err}	valid**
1	1	0.08	0.2	100%	0.27	0.54	97%
2	2	0.1	0.25	100%	0.31	0.66	94%
4	4	0.14	0.38	98%	0.55	1.06	75%
10	10	0.04	0.13	100%	0.19	0.34	96%
20	20	0.06	0.17	94%	0.07	0.20	81%

**) offsets and errors given in mm and deg   **) t_{err} <= 5mm, r_{err} <= 5 deg (RMSE)*

- Repeated movements with randomized initial and target pose.
- Projection-distortion on moving Stanford Bunny model is simulated. Pose offset is calculated from distortion in simulated camera image.

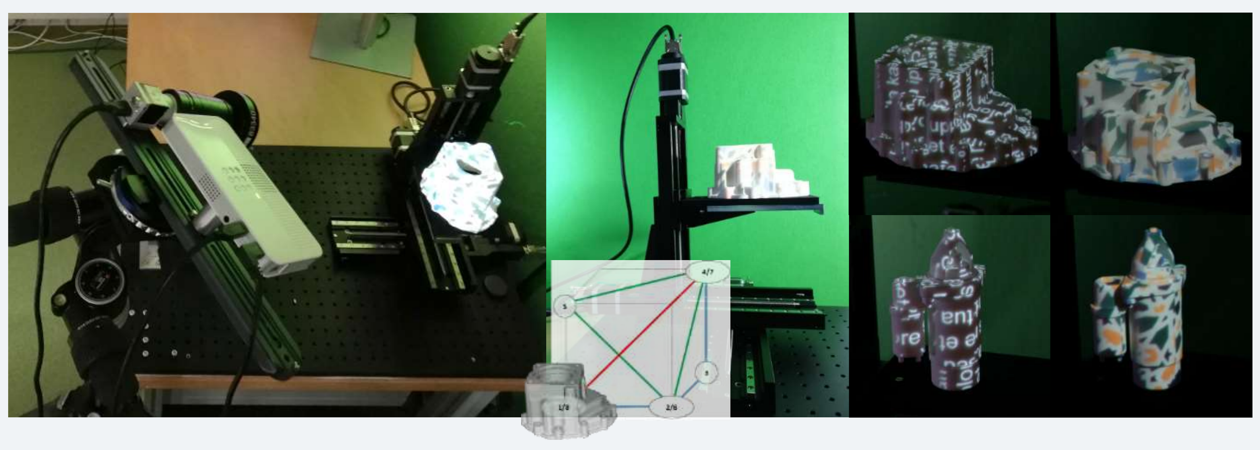
Real Data

- Measure deviation from ground truth path of linear translation unit.
- Two test objects and textures, four test cases:
  1. Projection-based tracking
  2. Projection-based tracking + ambient illumination
  3. Combined model + ambient illumination
  4. Ignore projection: contour + ambient illumination

	Obj. 1, Tex. 1*		Obj. 1, Tex. 2*	
Test	t_{err}	r_{err}	r_{err}	r_{err}
1	0.78	0.45	0.99	0.51
2	0.84	0.97	1.16	0.95
3	0.85	0.79	fails after Pos. 7	
4	fails after Pos. 4		fails after Pos. 4	

**) offsets and errors given in mm and deg*

- Freehand movements tested without ground truth comparison.



Reduction of aperture problem: Object 2 benefits from combined model.

(6) Conclusions

- **Pose estimation** of 3D objects by aligning simulated AR-projection with camera image.
- No need for additional hardware and radiometric calibration.
- Extension for extrinsic projector-camera **calibration.**

[1] N. Gard, P. Eisert, Markerless closed-loop projection plane tracking for mobile projector-camera systems, In *Proc. ICIP*, pp. 3363-3367. IEEE Computer Society, Washington DC, USA, 2018.
[2] E. G. Steinbach, P. Eisert, and B. Girod. Model-based 3-d shape and motion estimation using sliding textures. In *Proc. VMV*, pp. 375–382. AKA GmbH, Berlin, Germany, 2001.

Funded by: Federal Ministry of Education and Research
Supported by: Federal Ministry for Economic Affairs and Energy
on the basis of a decision by the German Bundestag

niklas.gard@hhi.fraunhofer.de
www.hhi.fraunhofer.de