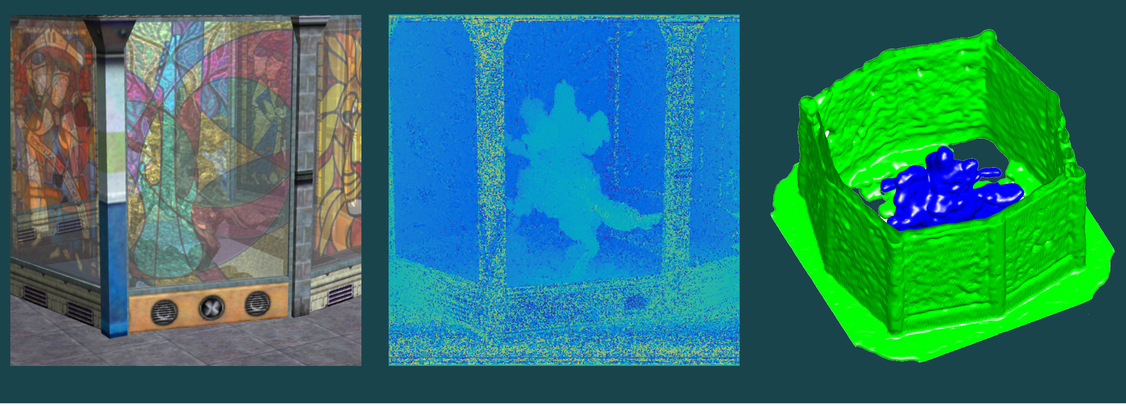
Layered Scene Reconstruction from Multiple Light Field Camera Views

Ole Johannsen, Antonin Sulc, Nico Marniok and Bastian Goldluecke

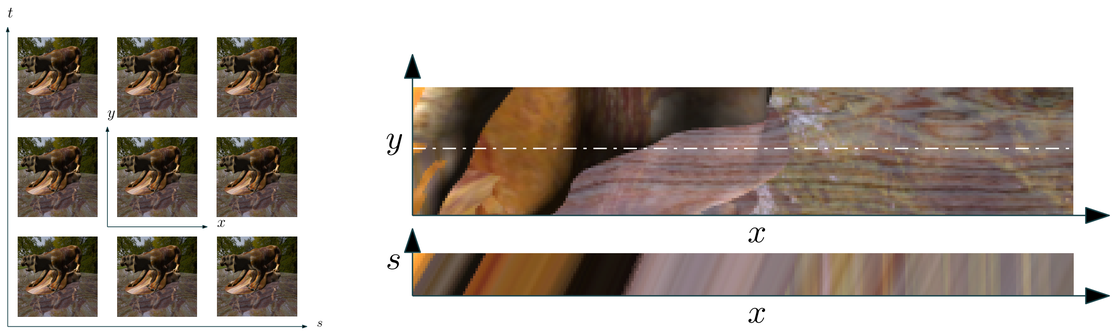
Computer Vision and Image Analysis,
University of Konstanz, {firstname.lastname}@uni-konstanz.de
The first two authors contributed equally.

Contribution

- A novel variational framework to infer geometry of scenes with semi-transparent or reflective materials.
- Light Fields camera views are leveraged to estimate multiple depth layers
- The different views are registered using a bundle adjustment framework specifically designed for light fields.



Light fields and Superimposed Layers



Light field: regular grid of views, identical cameras with parallel optical axes, parametrized with **view coordinates** (s, t) and **image coordinates** (x, y).

2D EPIs: horizontal slices with fixed (x, s) or vertical slices with fixed (t, y). Projection of a 3D point: line on an EPI, whose **orientation corresponds to disparity**.

Superimposed layers (reflections or tranparent objects): two super-imposed orientations which can be individually estimated [4, 2].

Pose estimation and Bundle Adjustment

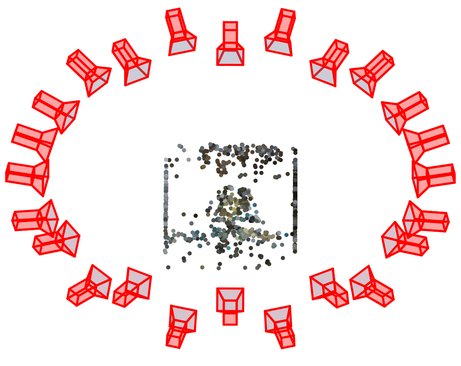
Structure from Motion

The projection of a single scene point/feature forms a 2D subspace of the 4D light field. Using this subspace as a matching criterion the relative pose between light fields can be infered [1].

Bundle adjustment

In this paper we use a non linear bundle adjustment to refine the initial esimate from [1] for multiple light fields. This increases the accuracy by up to a factor of 4.

		SfM		BA	
	Opacity	R error	t error	R error	t error
Temple	50%	49.723	48.7	**20.310**	**13.177**
	70%	137.204	121.523	**37.550**	**35.311**
	90%	97.783	72.764	**42.681**	**40.407**
Warrior	50%	64.615	66.867	**26.411**	**27.941**
	70%	68.501	80.288	**29.202**	**30.013**
	90%	15.358	22.238	**12.311**	**18.388**



Variational Surface Reconstruction

The goal is to segment the reconstruction volume $\Gamma \subset \mathbb{R}^3$ into two regions - the *object* Γ_{in} and the *empty* space Γ_{out}. The surface $\Sigma \subset \Gamma$ on the boundary of the two class regions corresponds to the reconstructed surface.

We use a variational framework similar to [3] to define a convex energy based on the theory of weighted minimal surfaces

$$E(u) = \lambda \int_\Gamma \rho \, |Du| + \int_\Gamma a u \, \mathrm{d}x \qquad (1)$$

where

$u : \Gamma \to \{0, 1\}$ is a binary indicator function,
$\rho : \Gamma \to \mathbb{R}^+$ is the surface consistency error and used as an weight for regularization (TV),
$a : \Gamma \to \mathbb{R}$ is a regional cost. $a < 0$ if a point has preference to be inside an object and $a > 0$ otherwise.

This work was supported by the ERC Starting Grant "Light Field Imaging and Analysis" (LIA 336978, FP7-2014), and the SFB Transregio 161 "Quantitative Methods for Visual Computing".

Presented at the Asian Conference on Computer Vision (ACCV), Taipei, November 2016.

Surface Consistency Error ρ and Region Cost a

Surface probability ρ and region cost α are updated iteratively with each view added to the reconstruction. For each view and every depth layer d, we compute a likelihood score c [2]. In addition, let $z(x)$ denote the distance to the camera for each point x in the reconstruction volume Γ. We distinguish three cases:

if $|z(x) - d(\pi(x))| \leq \epsilon$ the surfaces most likely runs throuth this voxel so we decrease the value of ρ with the respecive confidence c,

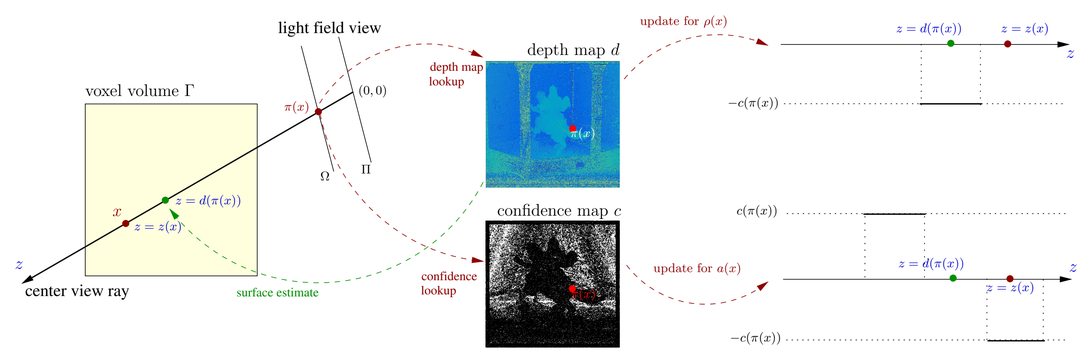
$$\rho(x) \leftarrow \rho(x) - c(\pi(x)),$$

if $d(\pi(x)) - \epsilon > z(x) \geq d(\pi(x)) - 2\epsilon$, i.e. the point is slightly in front of the estimated surface. The point is likely *outside* the surface and we increase the regional cost by the confidence score:

$$a(x) \leftarrow a(x) + c(\pi(x))$$

if $d(\pi(x)) + \epsilon < z(x) \leq d(\pi(x)) + 2\epsilon$, i.e. the point is slightly further away than the estimated surface, then the point is likely *inside* the surface and we decrease the regional cost by the confidence score:

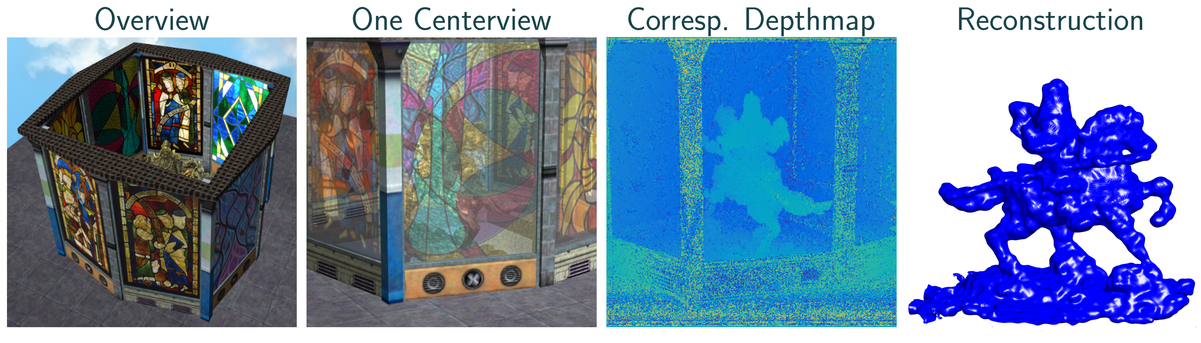
$$a(x) \leftarrow a(x) - c(\pi(x)).$$



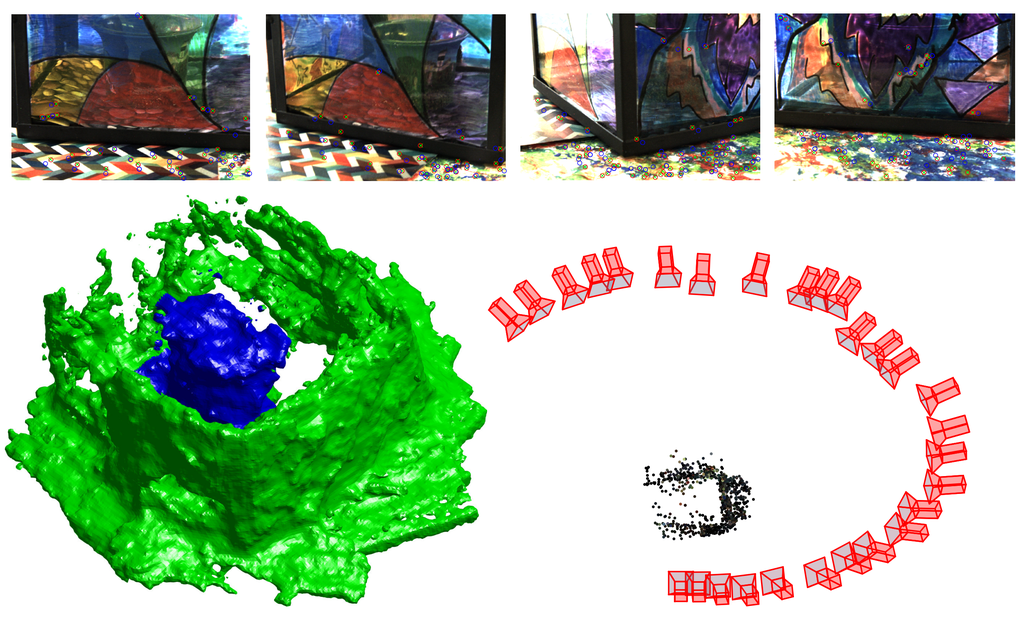
Results

Synthetic Example

Overview | One Centerview | Corresp. Depthmap | Reconstruction



Real-world Example (Lytro)



References

[1] O. Johannsen, A. Sulc, and B. Goldluecke. On linear structure from motion for light field cameras. In *Proc. International Conference on Computer Vision*, 2015.
[2] O. Johannsen, A. Sulc, and B. Goldluecke. What sparse light field coding reveals about scene structure. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3262–3270, 2016.
[3] K. Kolev, T. Pock, and D. Cremers. Anisotropic Minimal Surfaces Integrating Photoconsistency and Normal Information for Multiview Stereo. In *Proc. European Conference on Computer Vision*, 2010.
[4] S. Wanner and B. Goldluecke. Reconstructing reflective and transparent surfaces from epipolar plane images. In *German Conference on Pattern Recognition (Proc. GCPR)*, 2013.