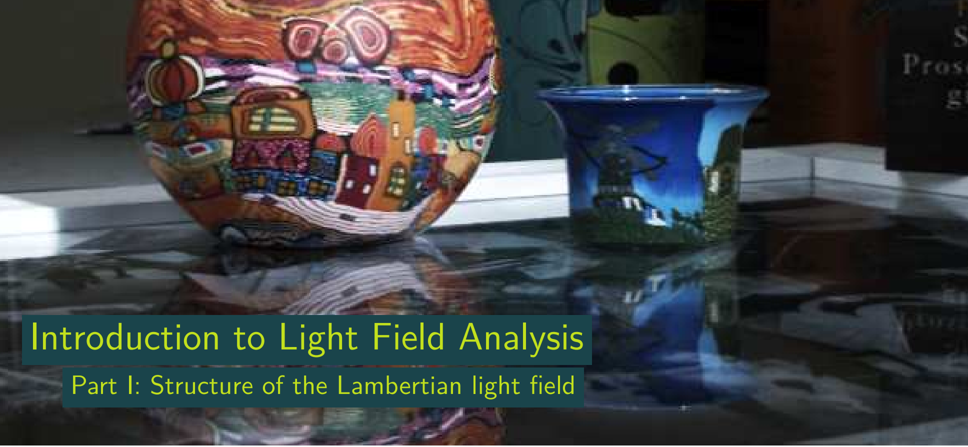


Introduction to Light Field Analysis

Part I: Structure of the Lambertian light field

Bastian Goldlücke and Ole Johannsen

GCPR Tutorial
October 2018

computer vision and image analysis

Universität Konstanz

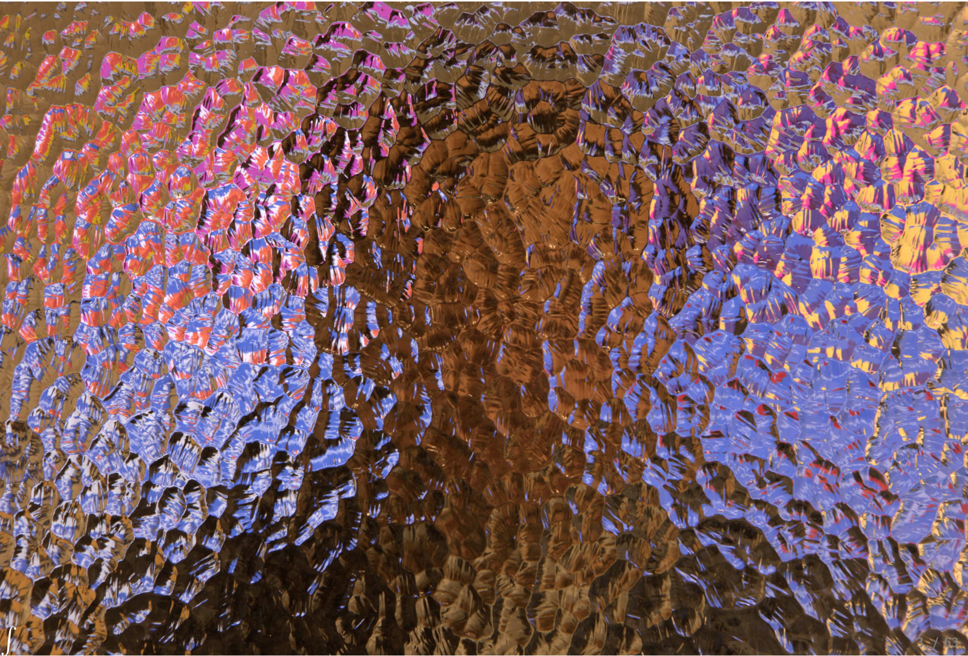
What is a light field?

The plenoptic function

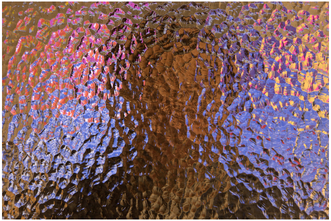
5D plenoptic function $L(x, y, z, \Theta, \Phi)$

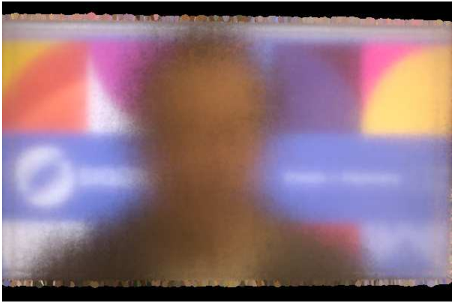
In computer vision usually only the "outside" 4D light field is considered

A pretty complicated “household” light field

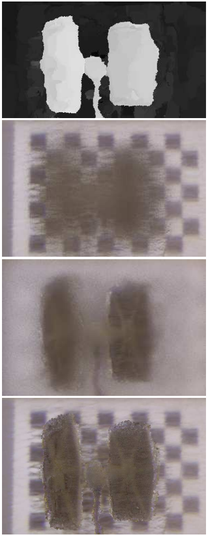


Resampling rays: synthetic aperture rendering





World's first spoon camera array

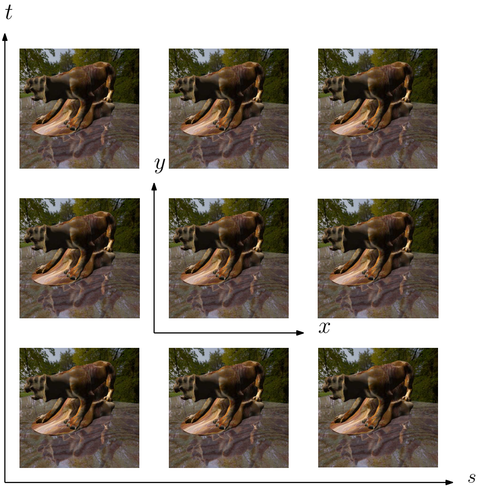


"The idea of this paper is insane." [Reviewer #2]

[Wender, Iseringhausen, Goldluecke, Fuchs, Hullin, VMV 2015]

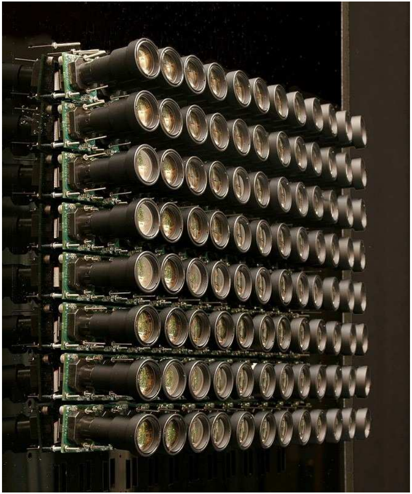
However we record a light field, for this tutorial,
we assume a representation in a simple standard structure.

A 4D lightfield for the purpose of this talk



Regular grid of **subaperture views**, identical pinhole cameras, parallel optical axes, parametrized with **view coordinates** (s, t) and **image coordinates** (x, y).

Capturing a light field



camera array

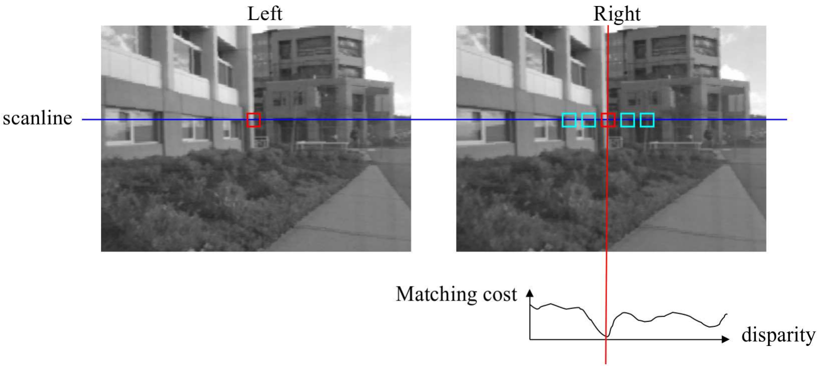
plenoptic camera

Key questions

- What is the structure of this representation, and what does a lightfield tell us about the 3D scene?
- How can we extend state-of-the-art image analysis techniques to light fields?

Quick reminder: two-frame stereo and cost volumes

Two-frame stereo and cost volumes



Disparity cost volume, e.g. pixel-wise

$$\rho(x, y, d) = \|I_L(x, y) - I_R(x - d, y)\|\,.$$

Many different (usually patch-based) cost-functions in use.

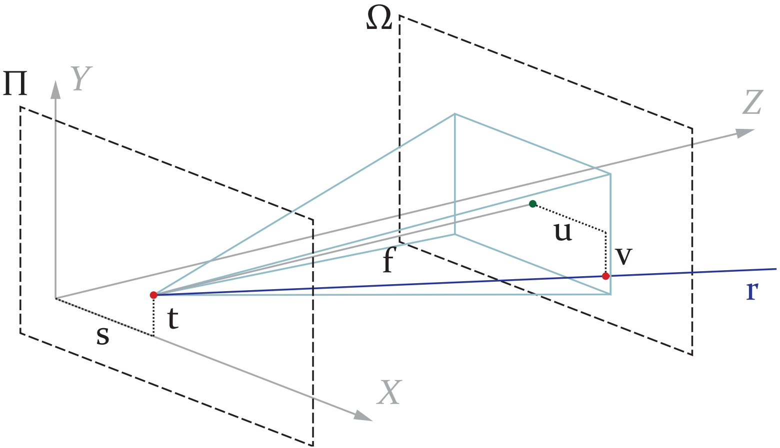
Point-wise costs and the need for optimization

- Often multiple local minima
- Flat regions: often no information, noise a problem
- Usual approach: embed cost function in global optimization scheme, e.g. solve

$$\operatorname*{argmin}_{u} \left\{ R(u) + \sum_{\boldsymbol{p}} \rho(\boldsymbol{p}, u(\boldsymbol{p})) \right\}$$

  with a regularizer R.
- Often spatially varying amount of regularization, depending on how much we trust the cost function.
- Some remarks on optimization later.

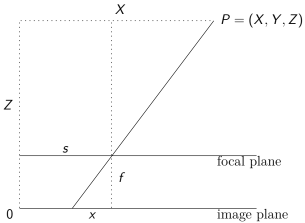
Light-field specific cost volumes?



Lightfield is a map on 4D space:

$$(x, y, s, t) \mapsto L(x, y, s, t) \qquad \text{or} \qquad (\boldsymbol{p}, \boldsymbol{b}) \mapsto L(\boldsymbol{p}, \boldsymbol{b})$$

with pixel coordinates $\boldsymbol{p}$ and camera coordinates $\boldsymbol{b}$.



Intercept theorem (pinhole perspective projection):

$$\frac{x}{f} = \frac{(X - s)}{Z}, \qquad \frac{y}{f} = \frac{Y - t}{Z}.$$

The projection coordinates for two different subaperture views (s_1, t_1), (s_2, t_2) satisfy

$$x_2 - x_1 = -\frac{f}{Z}(s_2 - s_1), \qquad y_2 - y_1 = -\frac{f}{Z}(t_2 - t_1).$$

Result: for a given depth (distance) Z of a scene point to the focal plane, there is a linear relationship between projection and view point coordinates. The "scale factor" $d = \frac{f}{Z}$ is called **disparity**.

Illustration: epipolar plane images

First idea: multi-frame stereo cost

- Compare pixel $\boldsymbol{p}$ in reference view I_R to corresponding pixel $\boldsymbol{p} - d\boldsymbol{b}_{V,R}$ in all others, i.e.

$$\rho(\boldsymbol{p}, d) = \sum_{V \neq R} \| \, I_R(\boldsymbol{p}) - I_V(\boldsymbol{p} - d\boldsymbol{b}_{V,R}) \, \| \,,$$

where $\boldsymbol{b}_{V,R}$ is the baseline between R and V.

- Straight-forward and works, but not very light-fieldish.

- Maybe main drawback: **No occlusion handling !**

Occlusion illustration

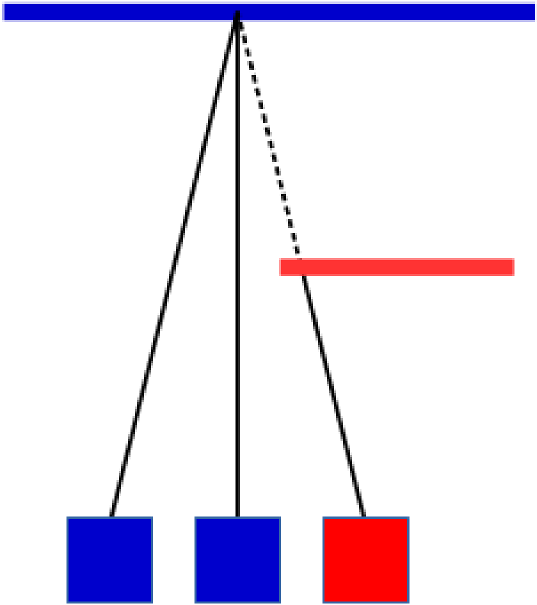
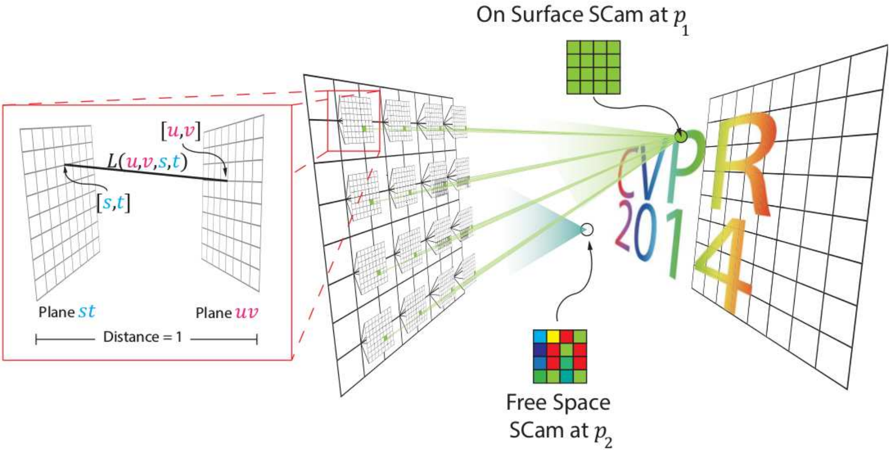


Figure: Wang et al. ICCV 2015

generalization: the surface camera (SCam)

SCam: projection of a 3D point into all LF views



Intuition: SCam is a camera at a 3D point looking at the light field planes.

Note: often, SCam views are called **angular patches**.

[Chen et al. CVPR 2014]

SCam or angular patch in formulas

- The **angular patch or SCam** $A_{\boldsymbol{p},d}$ for pixel $\boldsymbol{p}$ in the reference view and disparity d is

$$A_{\boldsymbol{p},d}(\boldsymbol{b}) = L(\boldsymbol{p} - d\boldsymbol{b}, \boldsymbol{b})$$

which depends on baseline $\boldsymbol{b}$. By convention, $\boldsymbol{b} = 0$ for the reference view (usually the center of the angular patch).

SCam or angular patch in formulas

- The **angular patch or SCam** $A_{\boldsymbol{p},d}$ for pixel $\boldsymbol{p}$ in the reference view and disparity d is

$$A_{\boldsymbol{p},d}(\boldsymbol{b}) = L(\boldsymbol{p} - d\boldsymbol{b}, \boldsymbol{b})$$

which depends on baseline $\boldsymbol{b}$. By convention, $\boldsymbol{b} = 0$ for the reference view (usually the center of the angular patch).

- Note: the standard stereo cost is a function of the corresponding angular patch,

$$\rho(\boldsymbol{p}, d) = \sum_{V \neq R} \left\| A_{\boldsymbol{p},d}(\boldsymbol{b}_{V,R}) - A_{\boldsymbol{p},d}(0) \right\| .$$

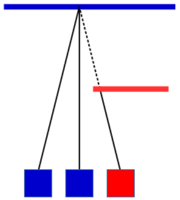
Another popular cost function is the variance of the angular patch, e.g. [Criminisi et al. 2005]

Best done for all pixels $\boldsymbol{p}$ in parallel:

- Choose disparity d
- Shift every subaperture view I_V by $d \cdot \boldsymbol{b}_{V,R}$ to align corresponding pixels with the reference view.
- The stack of transformed views T_V now corresponds to the SCam over every pixel.

Intuition: can be understood as “shearing” of the EPIs to make epipolar lines for a given disparity vertical.

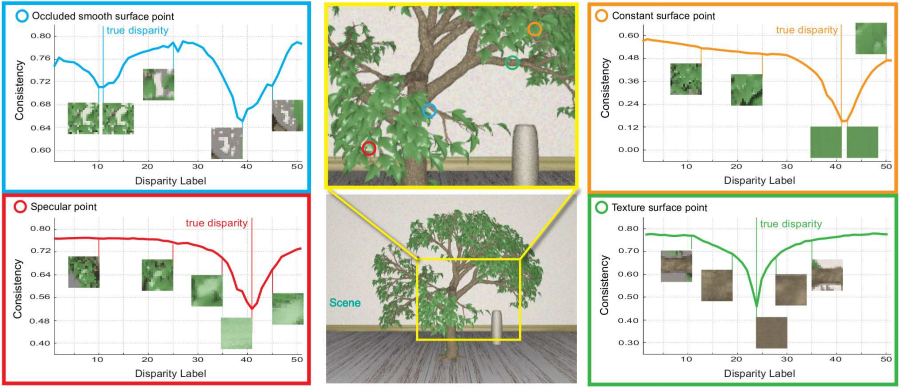
- Key idea: for a Lambertian unoccluded scene point, the SCam should be constant across all views.
- In empty space or inside an object, SCam pixels are probably inconsistent.
- At occlusion boundaries, there might be some pixels which are inconsistent, but one color should still dominate.



- Occlusion-aware cost intuition: Choose

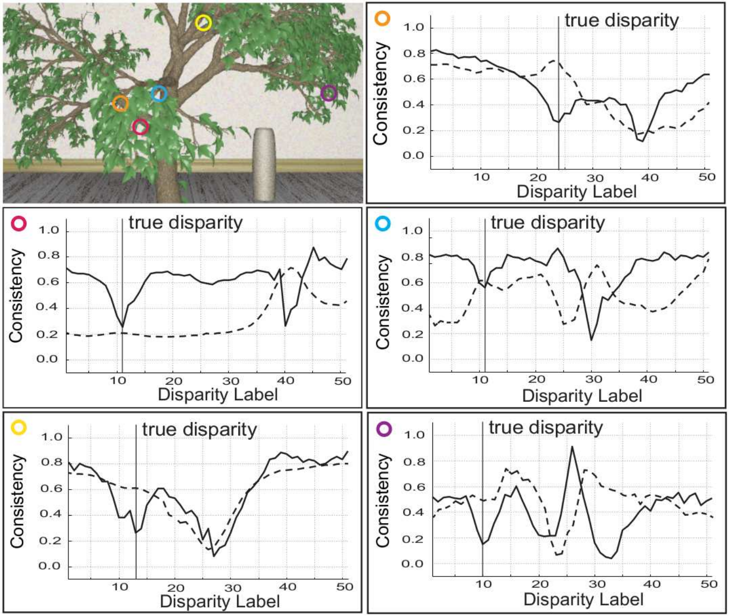
$$\rho(\boldsymbol{p}, d) = \begin{cases} \text{small} & \text{if } A_{\boldsymbol{p},d} \text{ contains a large low-variance region} \\ \text{large} & \text{otherwise.} \end{cases}$$

A possible implementation of this idea is [Chen et al. CVPR 2014].

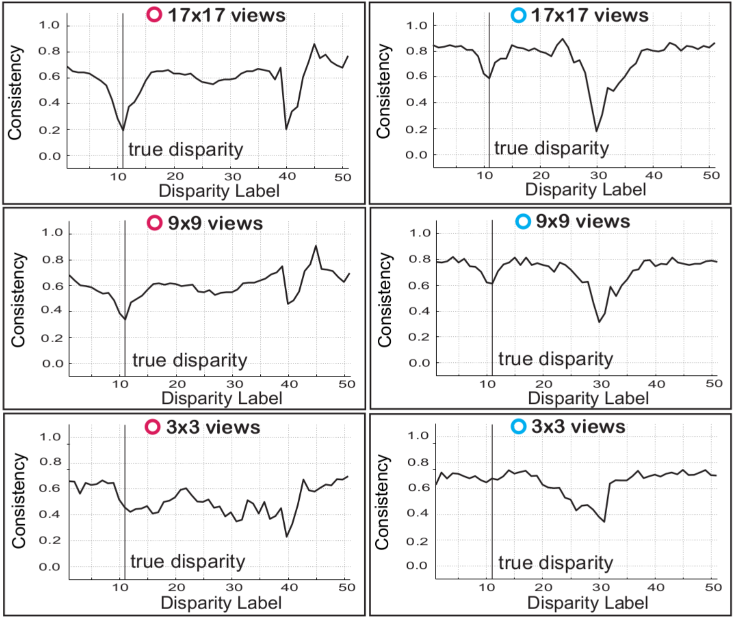


[Chen et al. CVPR 2014]

SCam vs. standard stereo dataterm



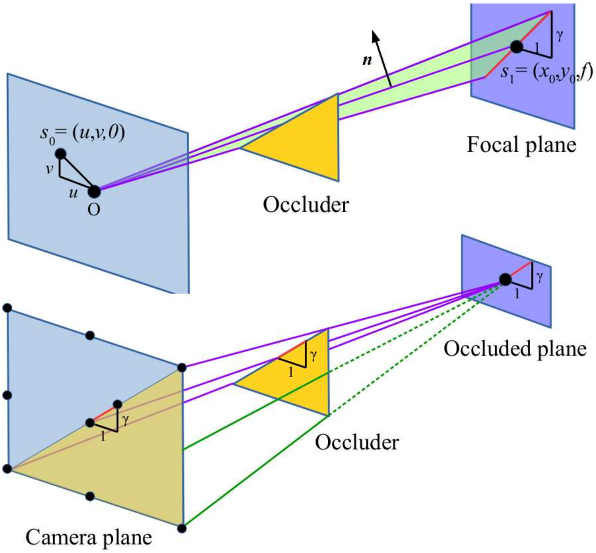
SCam view dependency



[Chen et al. CVPR 2014]

More sophisticated occlusion modeling

Occluder in angular patch



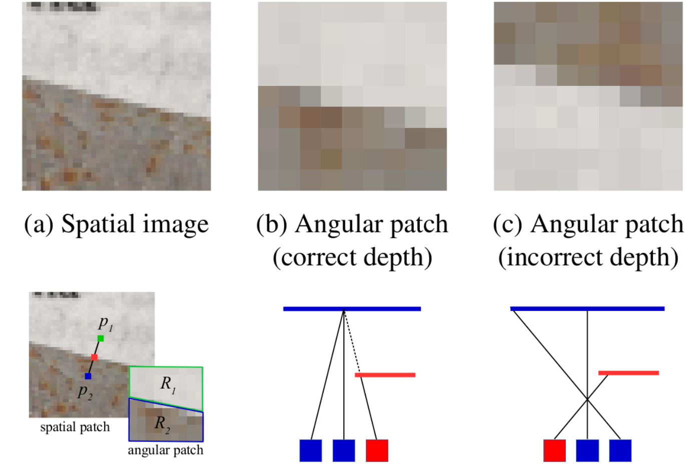
[Wang et al. 2015]

Intuition from the above illustration:

- Occluding edge orientation is the same in an angular patch as well as the center view.
- Thus, angular patch can be subdivided into occluded/unoccluded region with a single line parallel to the local image edge.
- The unoccluded region must have low color variance.

Occluder in angular patch: view point shift

[Wang et al. 2015]

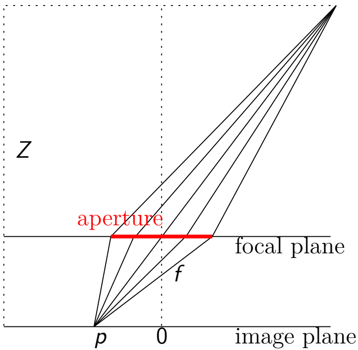


(a) Spatial image (b) Angular patch (correct depth) (c) Angular patch (incorrect depth)

(d) Color consistency (e) Focusing to correct depth (f) Focusing to incorrect depth

[Wang et al. 2015]

Depth from focus

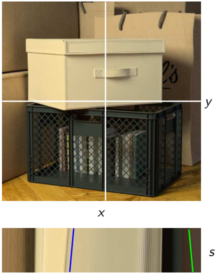


To construct refocused image at pixel $\boldsymbol{p}$ in the reference view, with camera focused at depth Z: sample over all rays in the subaperture views which correspond to $\boldsymbol{p}$.

$$F_Z(\boldsymbol{p}) = \sum_V w(V)\, L\left(\boldsymbol{p} - \frac{f}{Z}\boldsymbol{b}_V,\, \boldsymbol{b}_V\right).$$

The weight w describes e.g the virtual aperture, or other optical effects.

EPI view on refocusing



A light field is defined on a 4D volume parametrized by image coordinates (x, y) and view point coordinates (s, t). Epipolar images (EPIs) are the slices in the sx- or yt-planes depicted to the right and below the center view. By integrating the 4D volume along different orientations in the epipolar planes (blue and green), one obtains views with different focus planes.

Relation of refocusing to SCams

Refocusing can be formulated in terms of the angular patch corresponding to $\boldsymbol{p}$ and $d = \frac{f}{Z}$:

$$
\begin{aligned}
F_Z(\boldsymbol{p}) &= \sum_V w(V)\, L\left(\boldsymbol{p} - \frac{f}{Z}\boldsymbol{b}_V,\ \boldsymbol{b}_V\right) \\
&= \sum_V w(V)\, A_{\boldsymbol{p},d}(\boldsymbol{b}_V).
\end{aligned}
$$

Relation of refocusing to SCams

Refocusing can be formulated in terms of the angular patch corresponding to $\boldsymbol{p}$ and $d = \frac{f}{Z}$:

$$\begin{aligned} F_Z(\boldsymbol{p}) &= \sum_V w(V)\, L\left(\boldsymbol{p} - \frac{f}{Z}\boldsymbol{b}_V,\, \boldsymbol{b}_V\right) \\ &= \sum_V w(V)\, A_{\boldsymbol{p},d}(\boldsymbol{b}_V). \end{aligned}$$

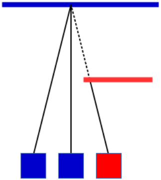
This shows that refocusing and depth reconstruction are intimately related. In particular, the pixel $\boldsymbol{p}$ is in focus if the angular patch $A_{\boldsymbol{p},d}$ has low variance.

Analogous to SCam computation:

- Choose disparity d
- Shift every subaperture view $I_{u,v}$ at (u, v) by $db_{u,v}$, where (u, v) is the baseline with respect to the reference view.
- The stack of transformed views $T_{u,v}$ corresponds to the SCam over every pixel.
- Compute weighted average over every pixel to generate refocused view.

Refocusing and occlusions

(b) Angular patch (correct depth)



(e) Focusing to correct depth

Even if focused at the correct depth, occlusions can lead to a blurred image as they "taint" the angular patches.

Focus measures and focal stack cost volume

Key idea: create a stack of images focused to different depths. The image from the stack which is "sharpest" around a pixel $\boldsymbol{p}$ corresponds to the correct disparity level.

Shape-from-focus: focal stack

Image: Agarwala et al. SIGGRAPH 2004

Shape-from-focus: focal stack

Image: Agarwala et al. SIGGRAPH 2004

Shape-from-focus: focal stack

Image: Agarwala et al. SIGGRAPH 2004

Shape-from-focus: focal stack

Image: Agarwala et al. SIGGRAPH 2004

Shape-from-focus: focal stack

Image: Agarwala et al. SIGGRAPH 2004

Shape-from-focus: focal stack

Image: Agarwala et al. SIGGRAPH 2004

Shape-from-focus: focal stack

Image: Agarwala et al. SIGGRAPH 2004

Shape-from-focus: focal stack

Image: Agarwala et al. SIGGRAPH 2004

Shape-from-focus: focal stack

Image: Agarwala et al. SIGGRAPH 2004

Shape-from-focus: focal stack

Image: Agarwala et al. SIGGRAPH 2004

Shape-from-focus: focal stack

Image: Agarwala et al. SIGGRAPH 2004

Shape-from-focus: focal stack

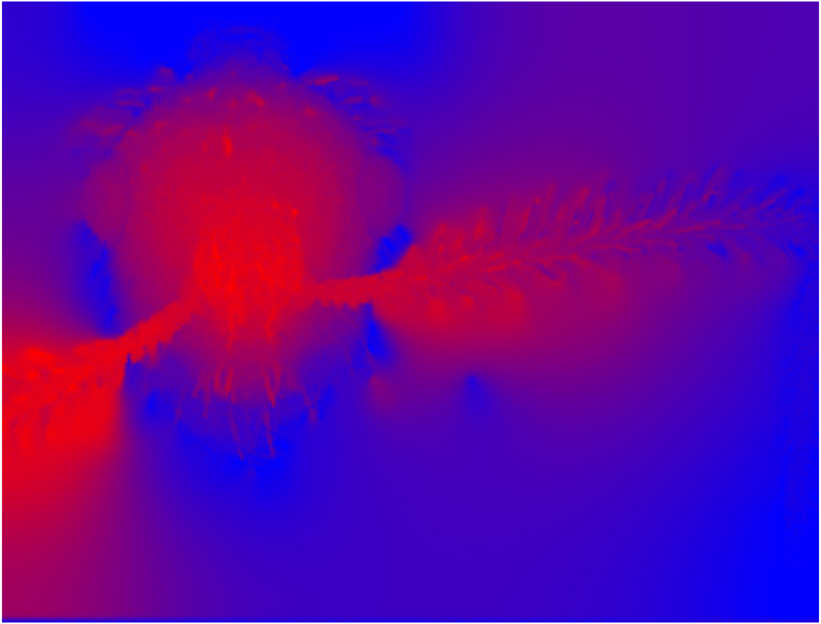
Image: Agarwala et al. SIGGRAPH 2004

- **Idea:** assign to each pixel in every image of the focal stack a number which tells us how sharp the surrounding image region is.

- Let $W(\boldsymbol{p})$ be a little window around the pixel $\boldsymbol{p}$ in image I_d focused at d, then a popular focus measure is the sum-modified Laplacian *[Nayar 1992]*

$$\rho(\boldsymbol{p}, d) = \sum_{\boldsymbol{q} \in W(\boldsymbol{p})} \left| \frac{\partial^2 I_d(\boldsymbol{q})}{\partial x^2} \right| + \left| \frac{\partial^2 I_d(\boldsymbol{q})}{\partial y^2} \right| .$$

Shape-from-focus: depth map

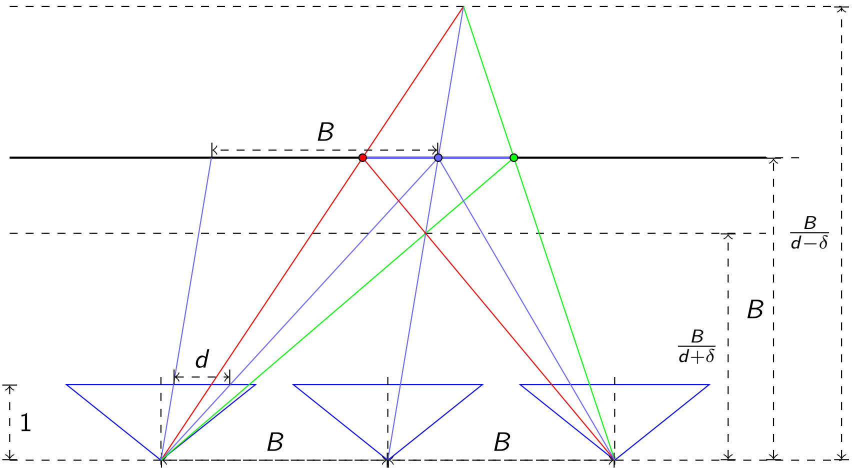


Focus stack / stereo integration

- There have been experiments which show that one can improve depth reconstruction by combining focus costs and stereo/SCam costs.
- However, it is not yet fully clear what the optimal weighting between those is (should be image-adaptive).
- In particular, where do we gain something from the focus measure which we cannot learn from the SCam directly?
- I believe a better idea is to use focal stack symmetry because this is more complementary, see next slides.

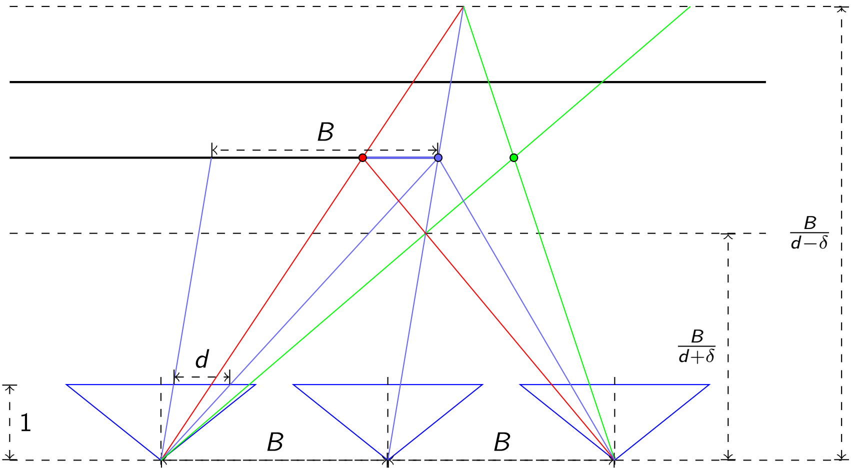
Focal stack symmetry

Focal stack symmetry



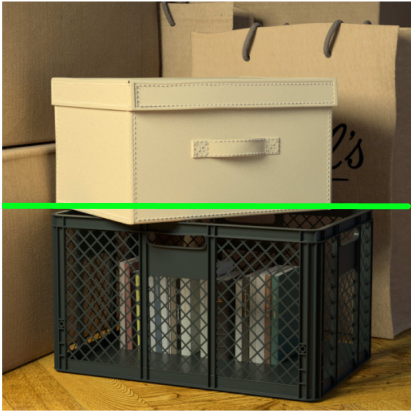
The focal stack is symmetric around the true disparity !

Focal stack symmetry

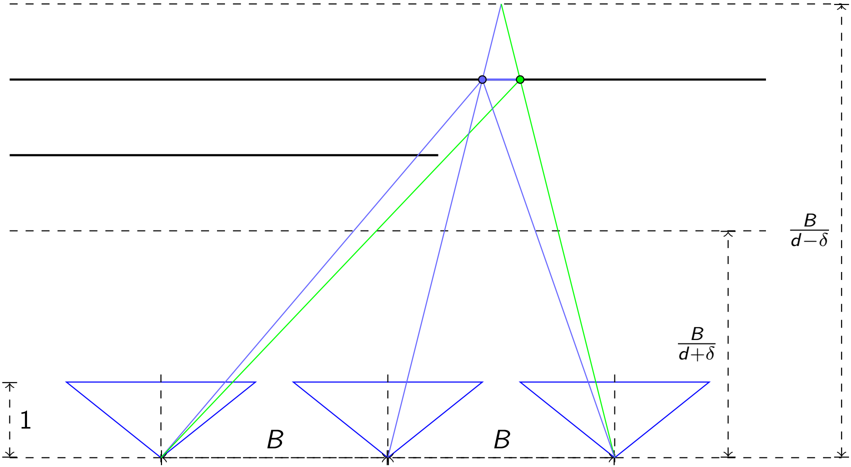


… however, occlusions destroy the symmetry property.

Focal stack symmetry

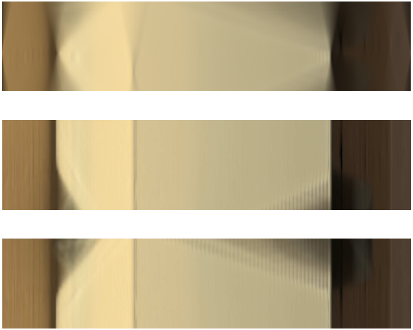


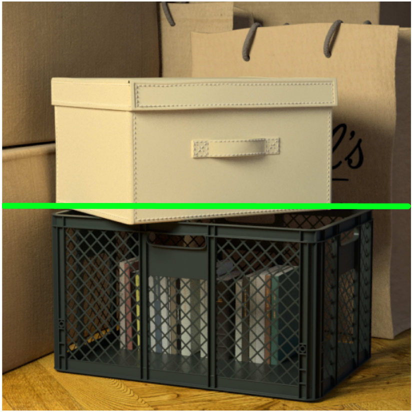
Remedy: partial focal stacks



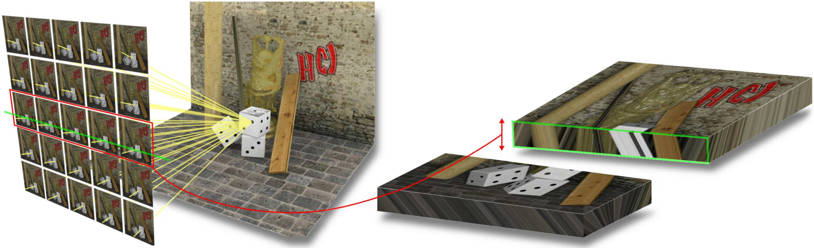
Under the assumption of not too small-scale occluders,
one direction is always occlusion-free.

Focal stack symmetry





Lambertian light fields: epipolar plane image structure



A 2D horizontal cut (green) is called an **epipolar plane image (EPI)**

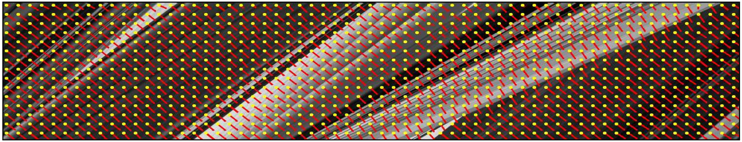
Wanner and Goldlücke, CVPR 2012 & TPAMI 2013

EPI from a recorded light field

[Wanner and Goldlücke, CVPR 2012 & TPAMI 2014]

Disparity estimation on an EPI

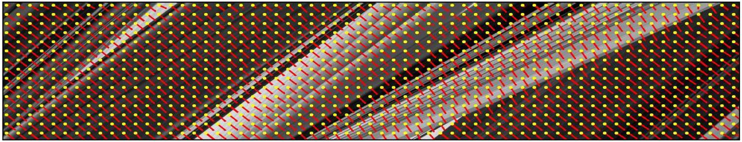
EPI from a recorded light field



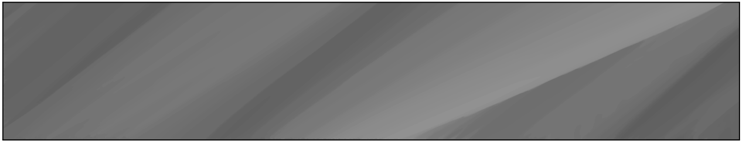
Structure tensor orientation estimate $\boldsymbol{e}_1(\mathcal{T}_{2.5})$

[Wanner and Goldlücke, CVPR 2012 & TPAMI 2014]

EPI from a recorded light field



Structure tensor orientation estimate $\boldsymbol{e}_1(\mathcal{T}_{2.5})$

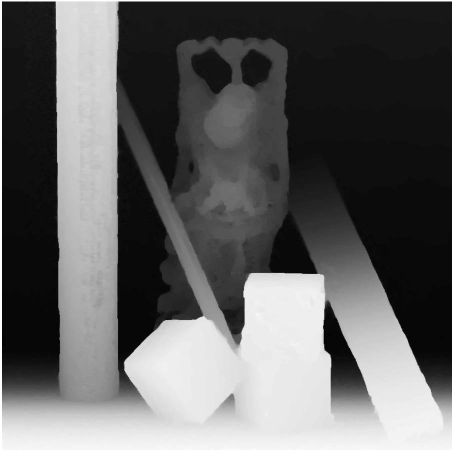


Resulting depth estimate (slope of orientation)

[Wanner and Goldlücke, CVPR 2012 & TPAMI 2014]

Dense depth via orientation estimation

light field center view



estimated depth map (two EPI orientations fused)

[Wanner and Goldluecke CVPR 2012, CVPR 2013, VMV 2013, TPAMI 2014]

EPI-orientation methods

Pattern/orientation analysis instead of matching.
Thus, perfect for CNNs

Benchmarks for disparity estimation

Old benchmark data set: HCI 2013

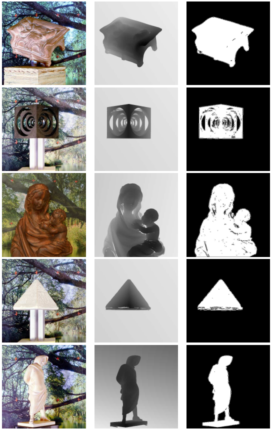
Custom-made benchmark for dense light fields

- 5 real-world and 7 synthetic datasets
- ground truth depth: Breuckmann smartSCAN

But

- insufficient accuracy of synthetic ground truth
- no centralized evaluation

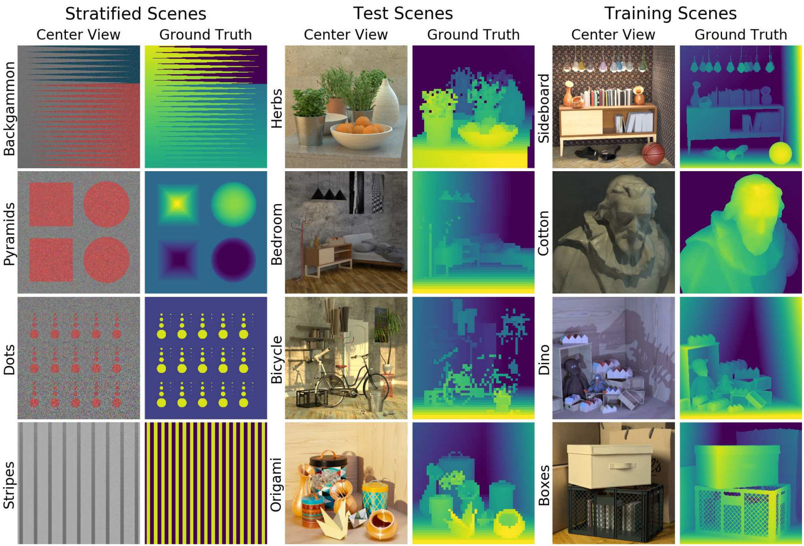
ray-traced light fields



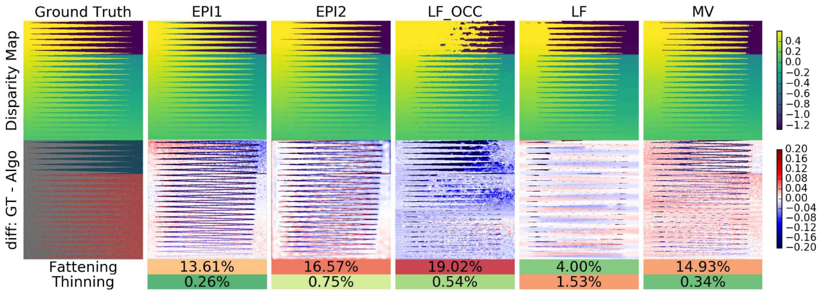
real-world light fields

Wanner, Meister and Goldlücke VMV 2013

A Benchmark for Depth Estimation on 4D Light Fields

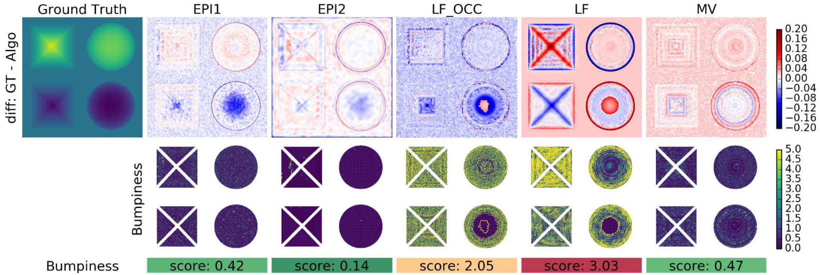


http://lightfield-analysis.net



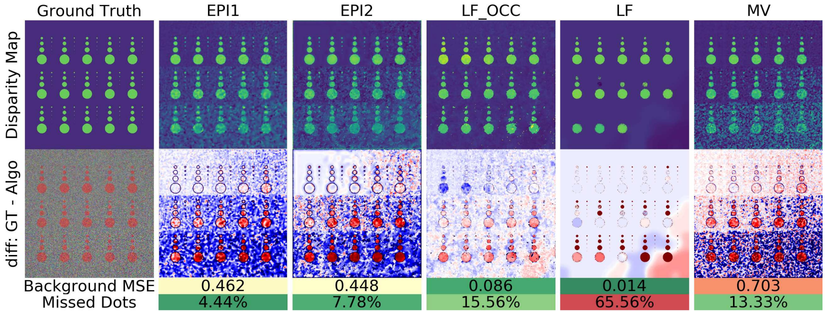
Backgammon

- Two slanted planes, zig-zag pattern cut out of foreground plane.
- Evaluates occlusion performance for varying distances between foreground and background.
- Specific contrary metrics: fattening and thinning.



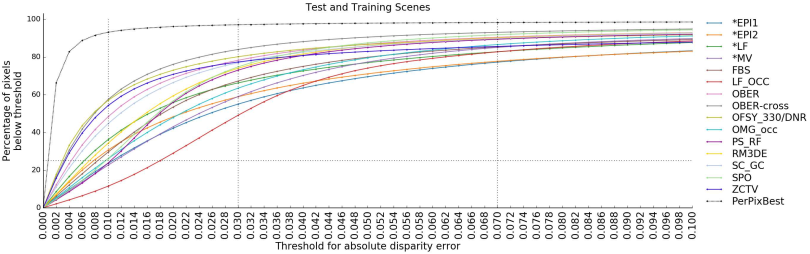
Pyramids

- Pyramids and hemispheres ontop/cut out of block.
- Evaluates smoothing capabilities, estimation of surface normals.
- Specific metrics: bumpiness (and mean angular error).



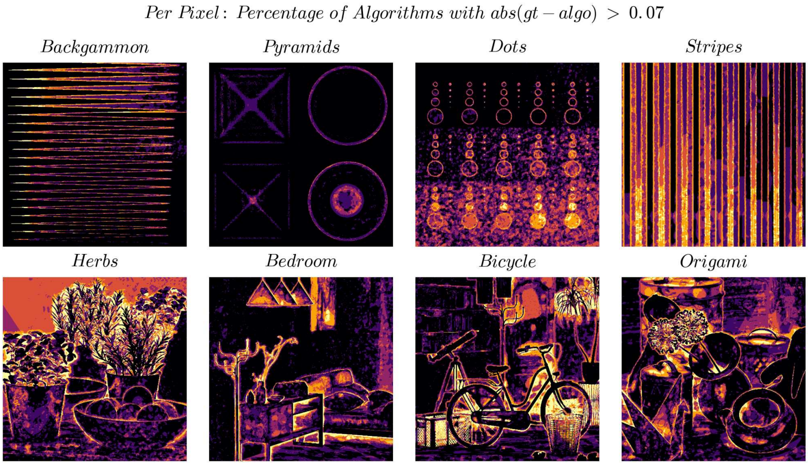
Dots

- Repeated pattern of dots of different sizes with inceasing amounts of noise.
- Evaluates performance under influence of noise.
- Specific metrics: background MSE and missed dots.



- Percentage of good pixels for different thresholds.
- Different maximum accuracy for different algorithms.

A Benchmark for Depth Estimation on 4D Light Fields



- Which parts of a scene are challenging?
- Occlusion areas, as corresponding to metric.

Summary

Summary

- Disparity and depth reconstruction
  - SCams and angular patches
  - Angular patch consistency
  - Occlusion modeling
  - Refocusing and focal stacks
  - Focal stack symmetry
  - Epipolar plane image structure
- Benchmarking
  - Specific scenes and metrics
  - Quality is ambiguous