SCI | STANFORD COMPUTATIONAL IMAGING

LiFF: Hand-Crafted Light Field Features

Donald G. Dansereau
SCI Group Meeting
09 Feb 2018

Outline

Motivation: Why LF Features?

LF Structure Review

Previous Approaches

Hand-crafted LF Feature

Preliminary Results

Dataset

Feature Detection & Matching

...the basis for much of computer vision

[COLMAP tutorial]

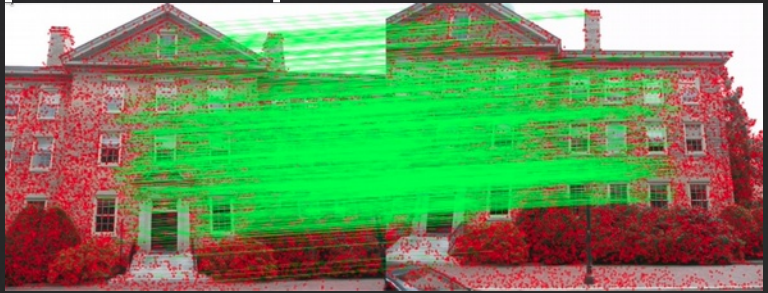


Image registration
Pose estimation
3D reconstruction
Place recognition
Change detection
…

Detect → describe → match → outlier reject & register

"It works pretty well most of the time"

→ Survivor bias

We don't use it where it doesn't work

Frequent failures in challenging conditions

Computation vs robustness tradeoff

Infrequent failures can kill

Where it Breaks

Low light / low contrast

Occlusions

Speculars

Reflection

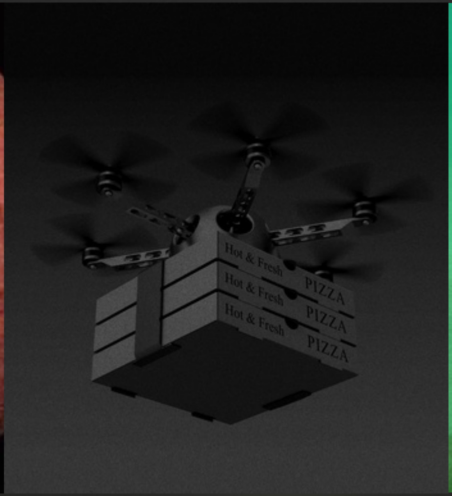
Transparency

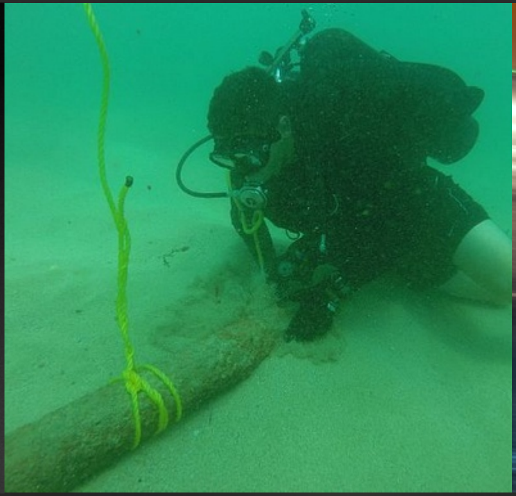
Driving, Drones

Delivery, surveillance, monitoring

Augmented reality

Medical, underwater, space, ...





LF Features

Low light / low contrast

Occlusions

Speculars

Reflection

Transparency

↑

Light fields are good at these

Detect → describe → match → outlier reject & register

LF Features

More selective detections

More robust and informative (3D) descriptors

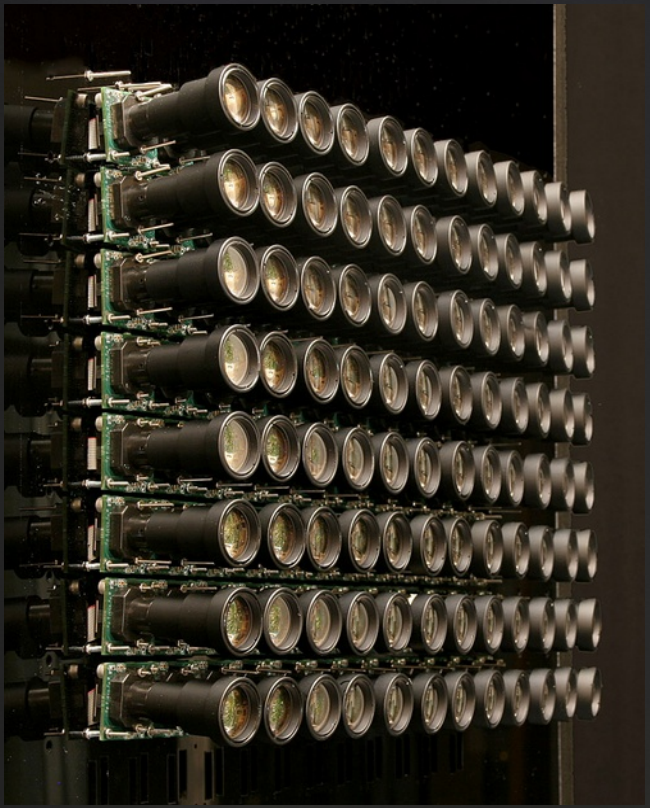
More selective matches

→ Fewer missed and incorrect registrations

New application areas

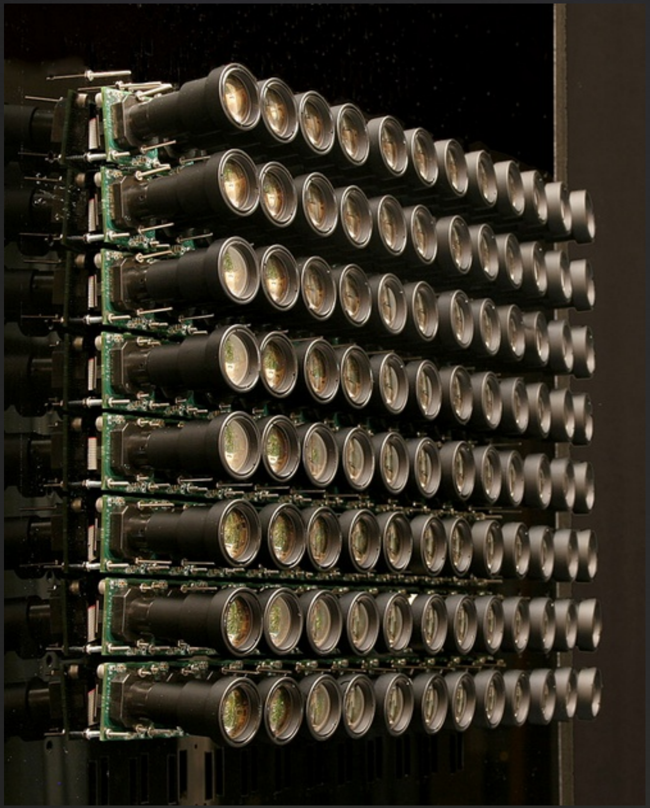
Saved dollars and lives

Review: LF Structure



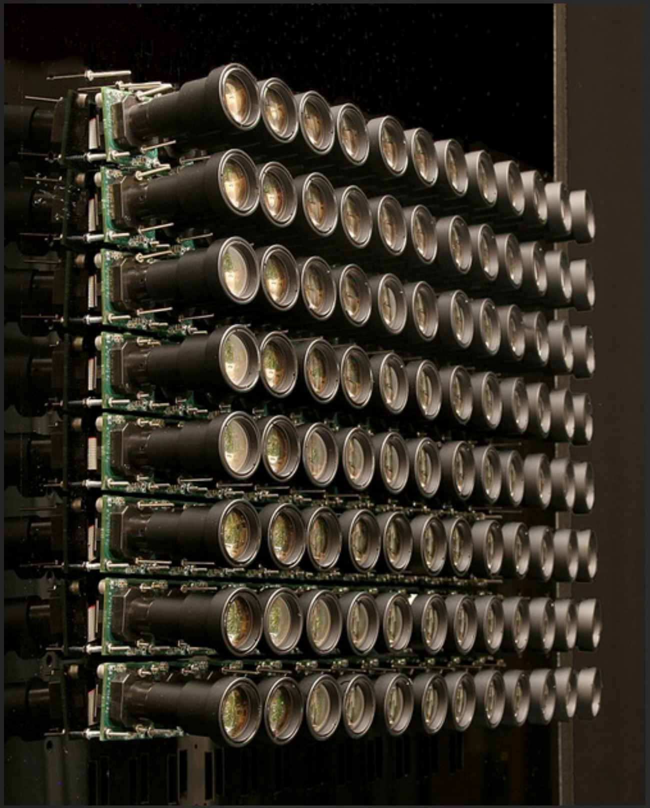
Stanford camera array

LF Structure



Stanford camera array

LF Structure



Stanford camera array

LF Structure

Lytro Lenslet-based camera

LF Structure

Lytro Lenslet-based camera

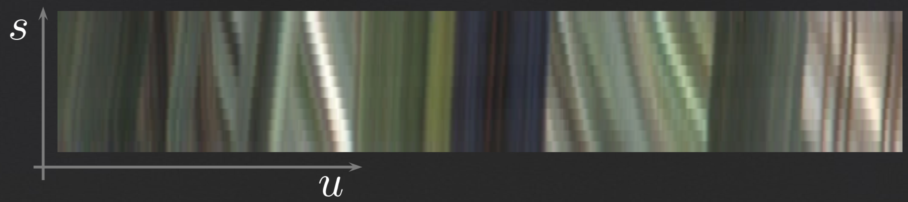
LF Structure

Lytro Lenslet-based camera

LF Structure

Lytro Lenslet-based camera

Epipolar Images



Previous Work: Hand-Crafted LF Features

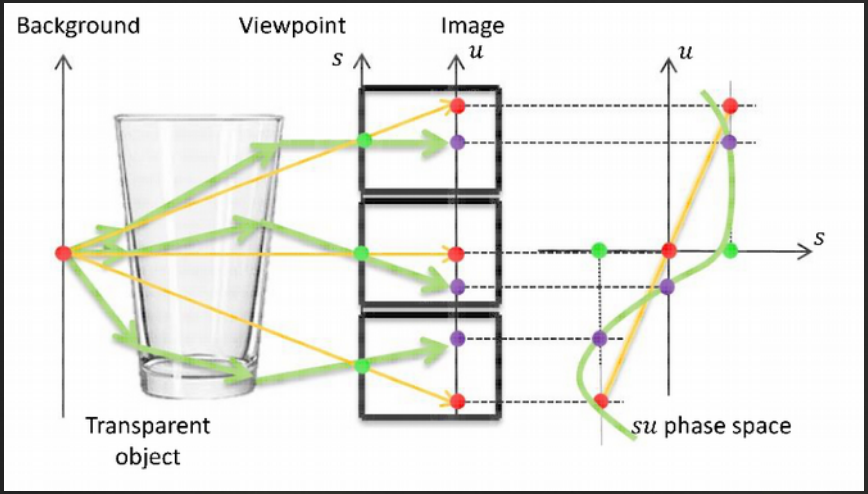
[Teixeira 2017 “3D keypoint detection by light field scale-depth space analysis”]

→ 2D SIFT in all sub-views, Hough to find lines in EPI images

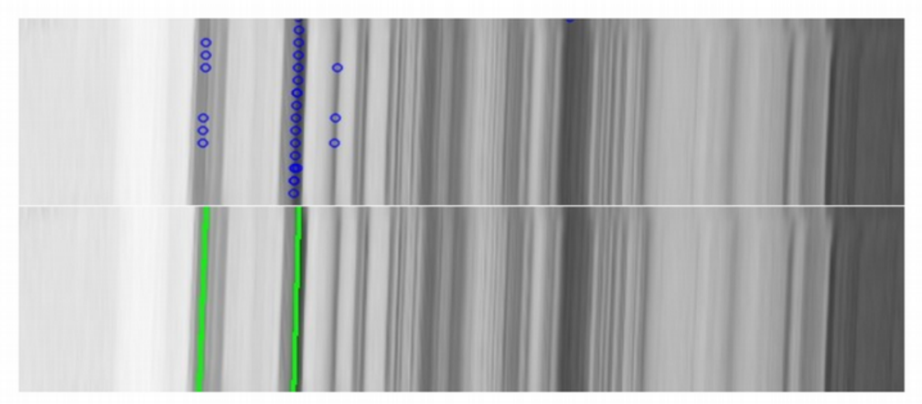
→ >200x slowdown; 2D decisions ignore 4D structure

[Xu 2015 “Transcut: transparent object segmentation from a light-field image”]

→ Optical flow between sub-views then a (partial) planarity check to find refractions



[Xu2015]



[Teixeira2017]

Opportunity and Approach

Previous work doesn't exploit LF structure for speed or robustness

An ideal feature

- Is fast
- Is robust
- Just works (with minimal tuning)

1) Learning LF Features (ongoing @Stanford)

2) Fast refractive feature rejection (ongoing w/QUT)

3) A hand-crafted generalization of SIFT

Contributions

1) LF feature detector and descriptor

    Faster than current approaches (typically 20x faster)

    More robust in low light, occlusions, non-Lambertian, ...

2) Quality filter and disparity estimate

    Additional robustness to refractions, reflections, occlusions, …

    Augmenting descriptor with disparity (depth)

[3) Adaptive operation]

    Exploit favourable conditions

    Estimate: additional 3x speed boost

4) Multiview LF Dataset (name?)

Review: SIFT

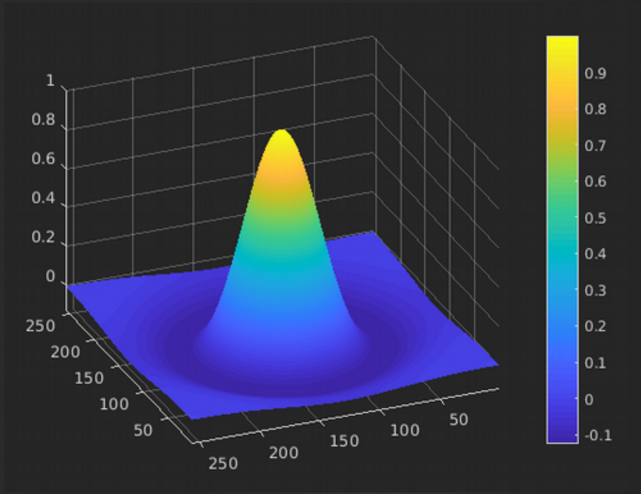
N_f x [patch]

2D ➡ **Detector** $\xrightarrow{N_f \text{ x } [u,v]}$ Descriptor ⟶ N_f x [u,v,**h**]

Detector: Difference of Gaussians

N_{scales} ⟶

2D $*$  …



… then find local extrema.

State-of-the-Art: Repeating 2D SIFT

… then combine (ad-hoc)

No benefit in low light

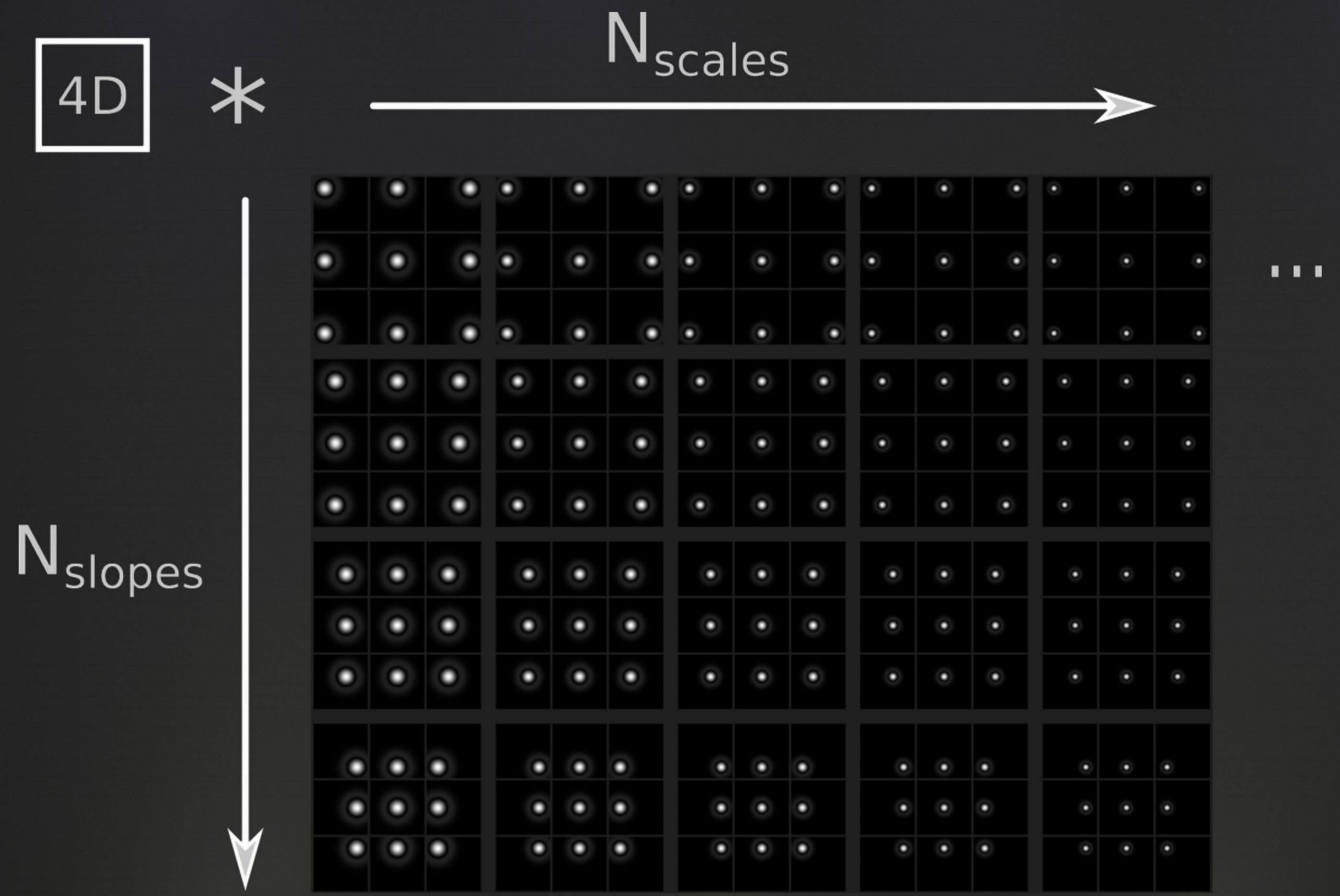
Descriptor sensitive to occlusions

Sensitive to non-Lambertian surf.

Slow! (Ns x Nt x slower)

[Teixeira 2017]

Full 4D: Jointly Detecting Scale, Slope



See through occlusions

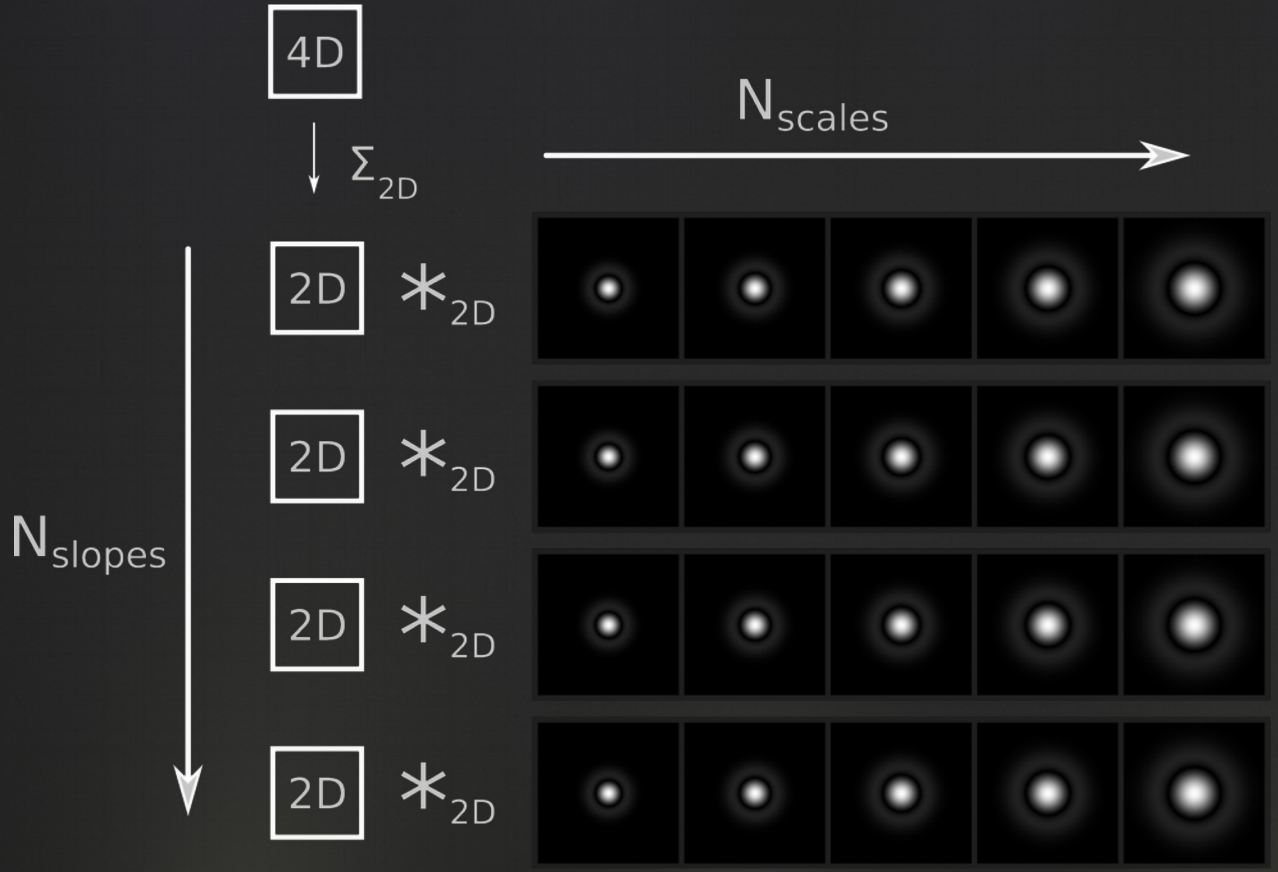
See in low light

Detect non-Lambertian

Est. depth (slope)

VERY slow

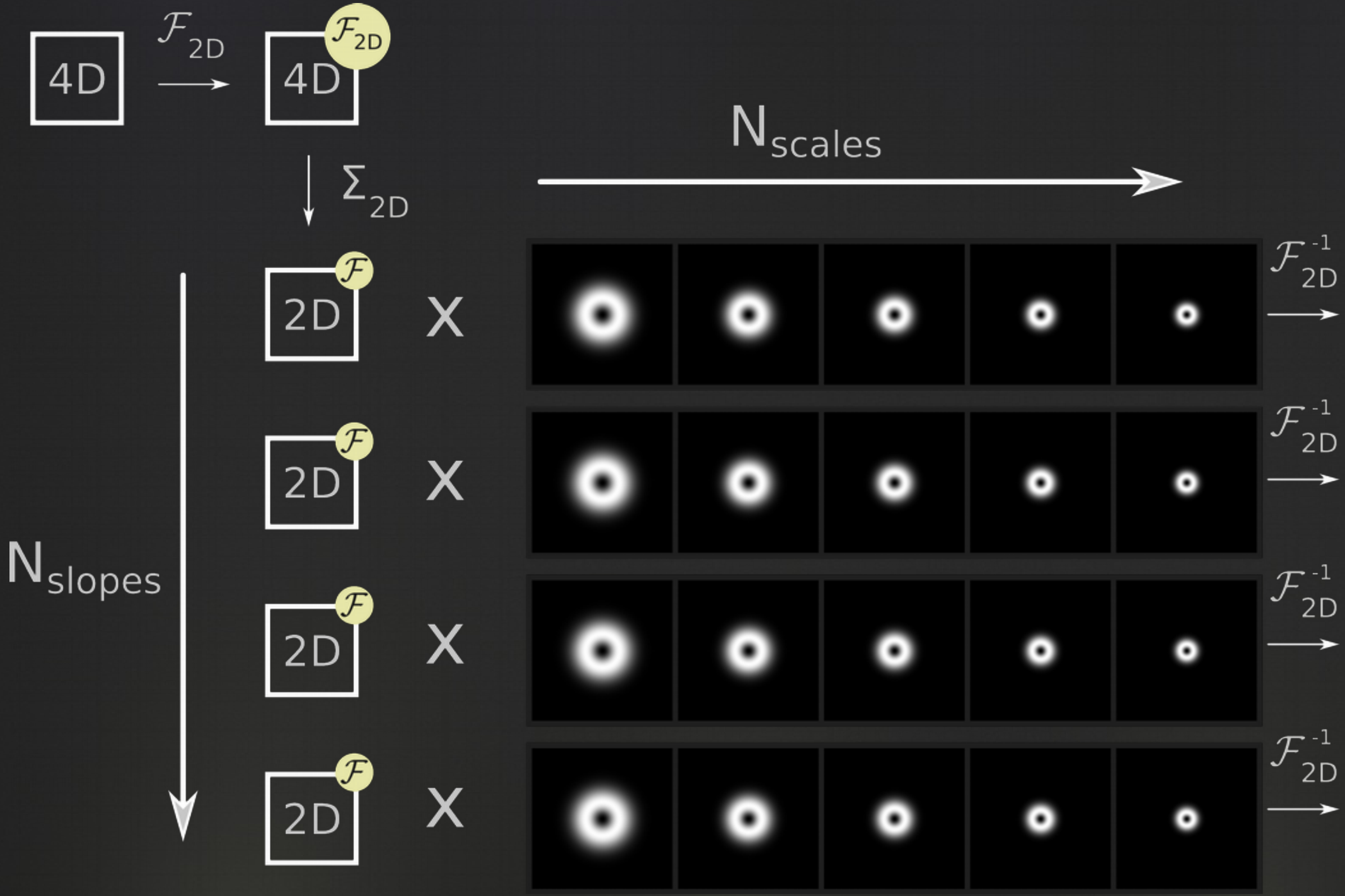
Separability → Speed



Identical to Full 4D

Much faster

Mixed-Domain Filtering

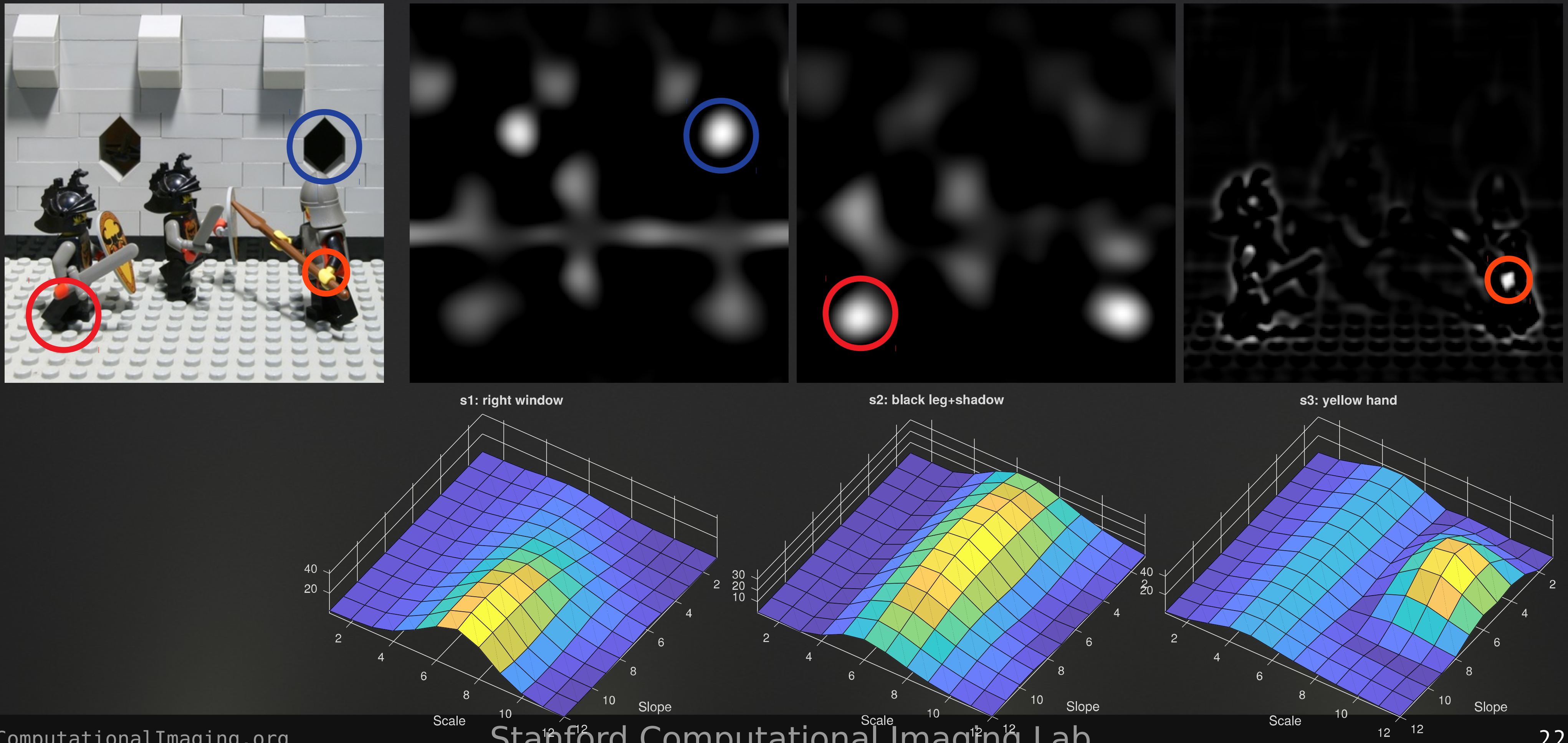


Faster in some scenarios

Depth *volumes*

→ Fewer slices

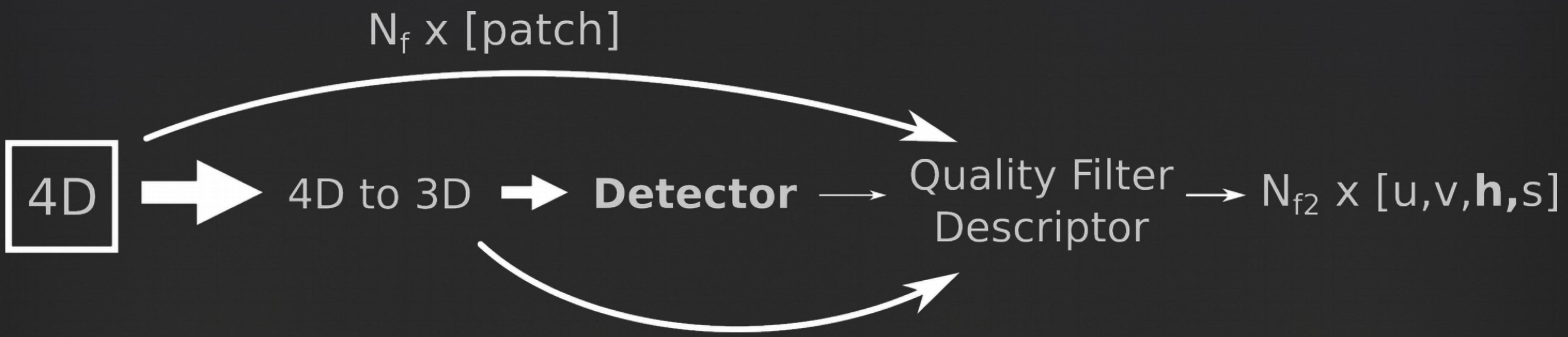
Demo of Joint Slope / Scale Estimation



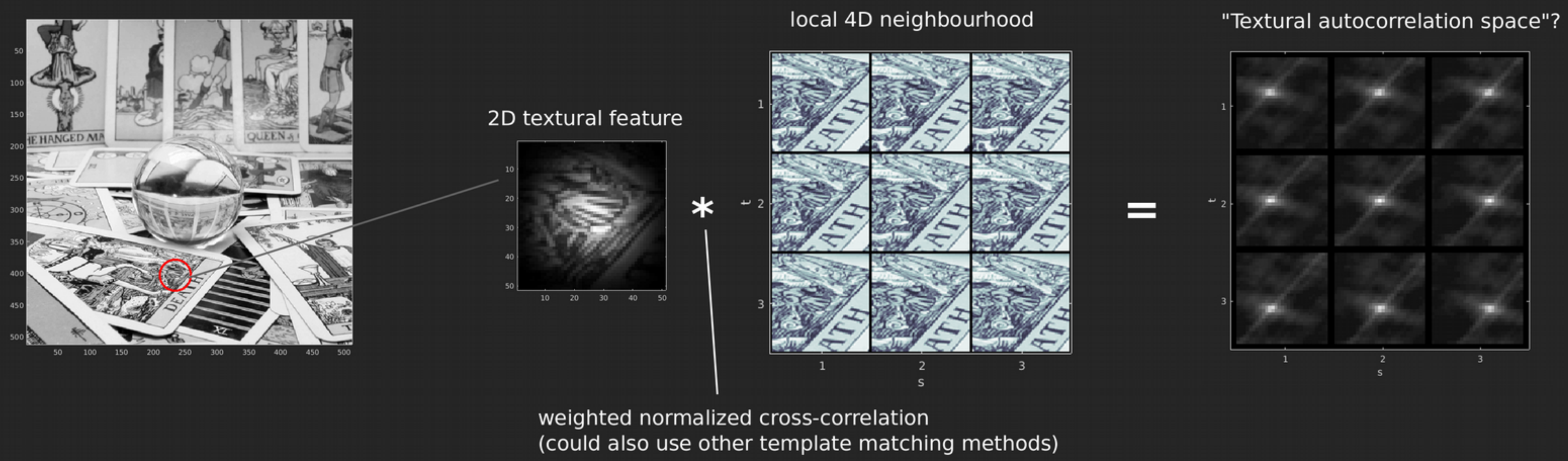
Complexity

Method	Est. Relative Speed (bigger is better) Illum, 16 scales, 8 slopes	Est. Relative Speed (bigger is better) Gantry, 16 scales, 8 slopes
Naive SIFT 4D	1	1
4D Full	1 / 21	1 / 21
2D FFT	9.8	11
2D Spatial	**22**	**27**
Adaptive	Up to 60	Up to 100

Quality Filter



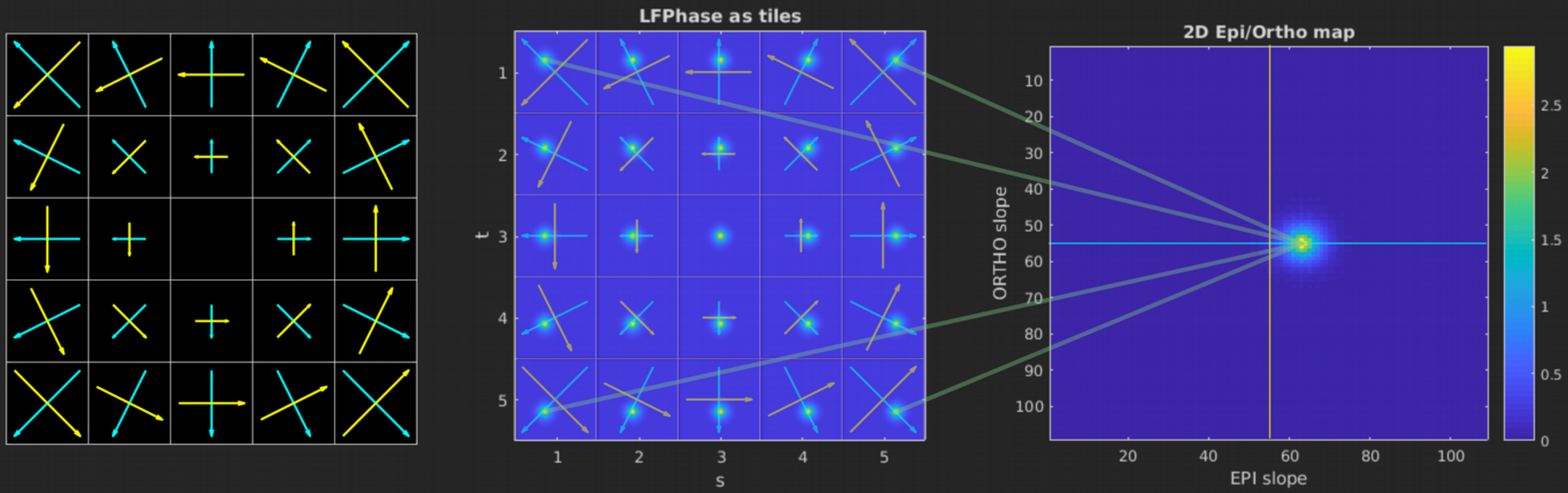
Local 4D Structure from Autocorrelation



Abstracts away from texture

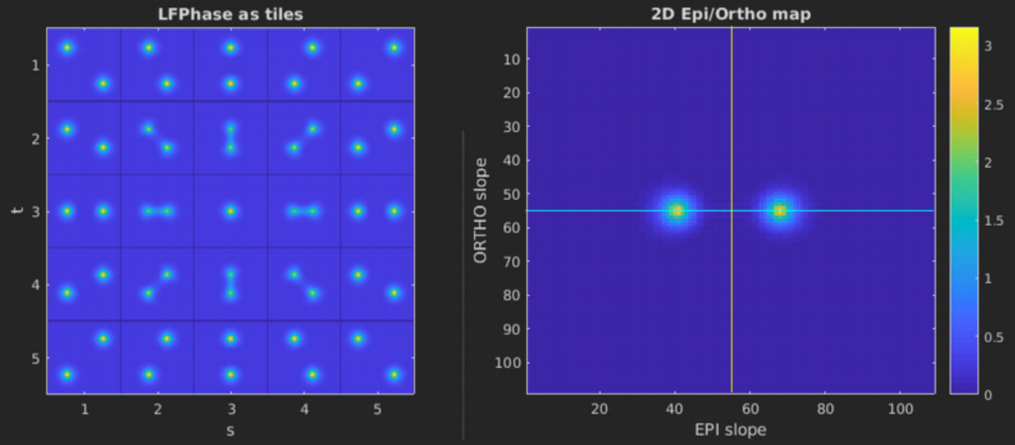
Captures non-lambertian effects

Epipolar Voting: 4D → 2D

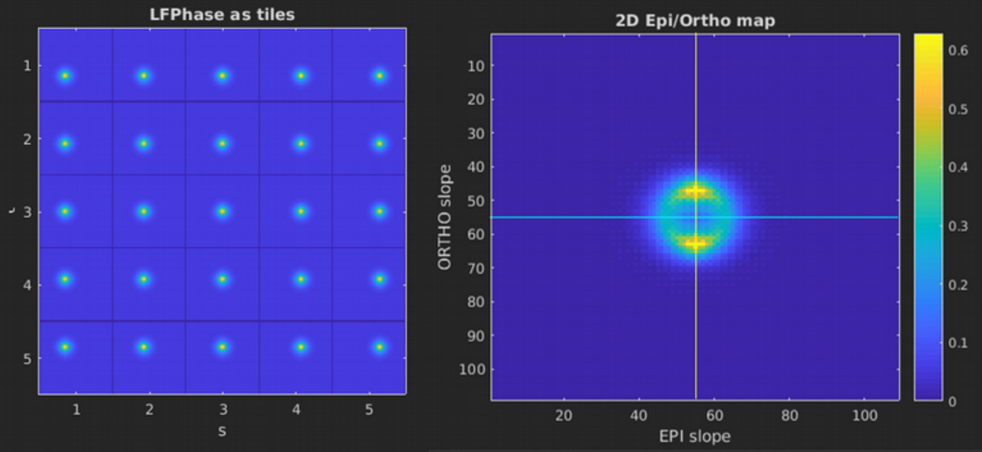


Reflections, Refractions

Scene with occlusion or reflection



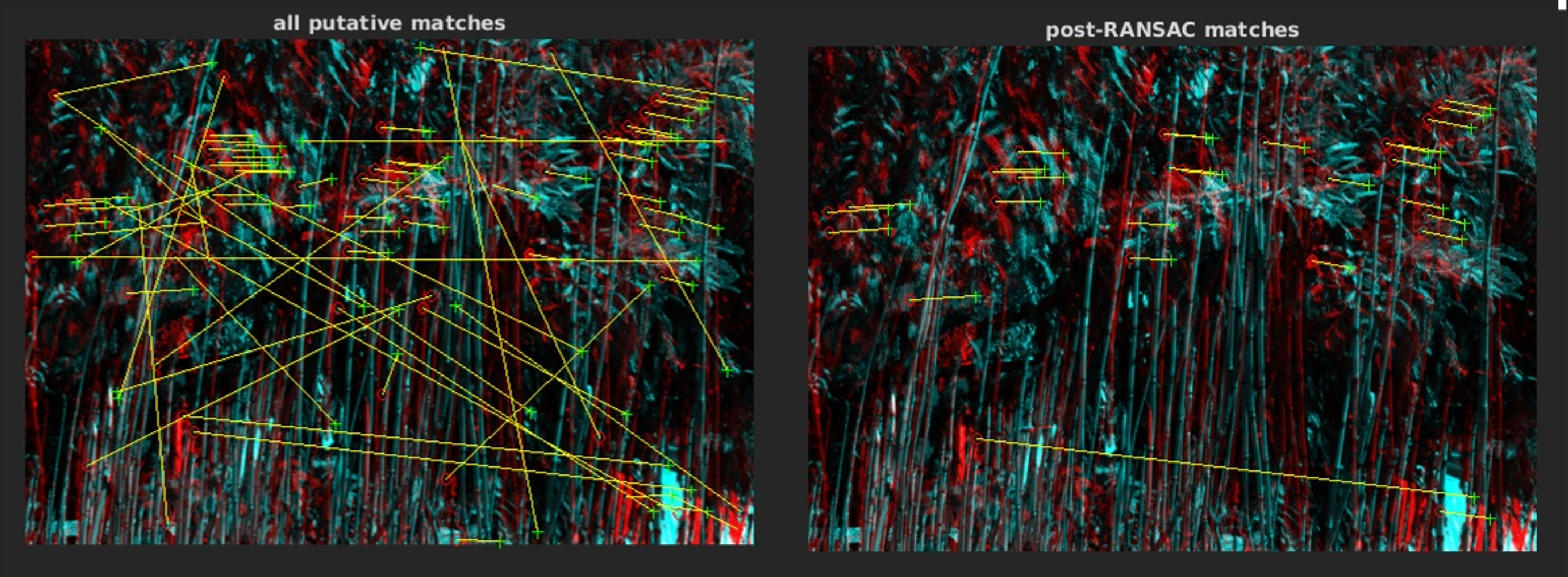
Cylindrical Refraction



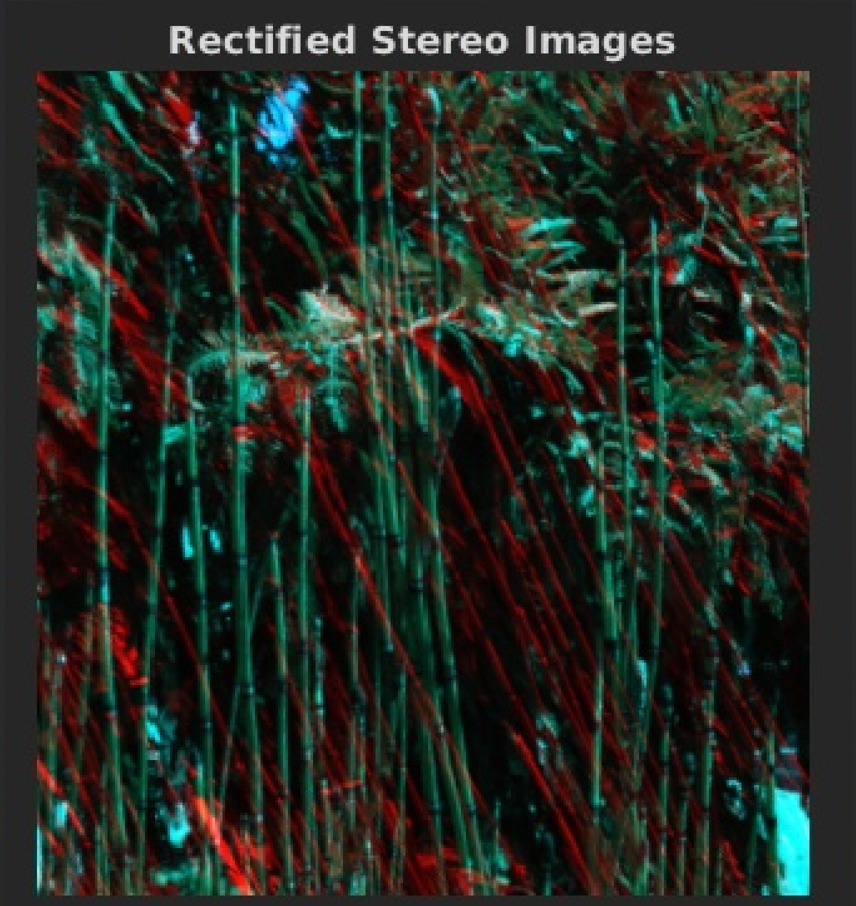
Proof of Concept: Focal stack feature

2 Lytro Illum images treated in 2D with stereo registration

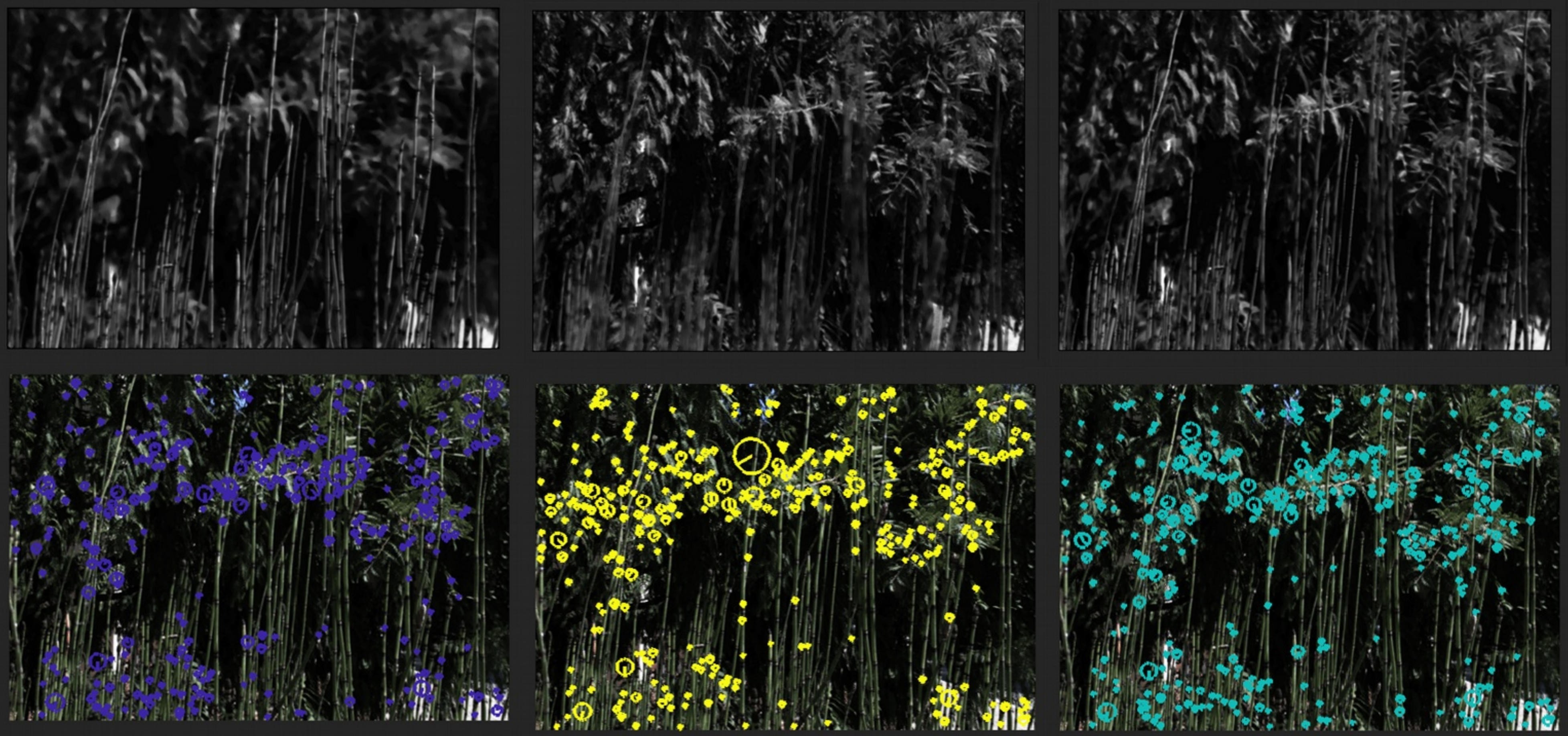
Proof of Concept: Focal stack feature



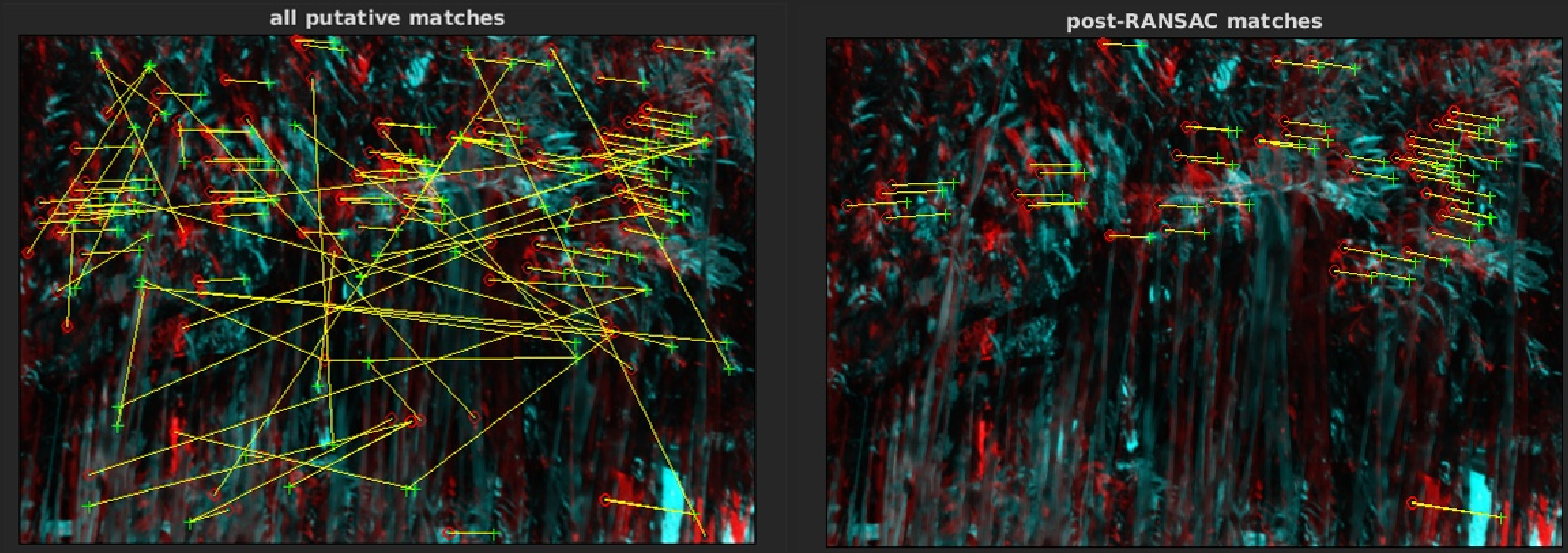
Proof of Concept: Focal stack feature



Proof of Concept: Focal stack feature

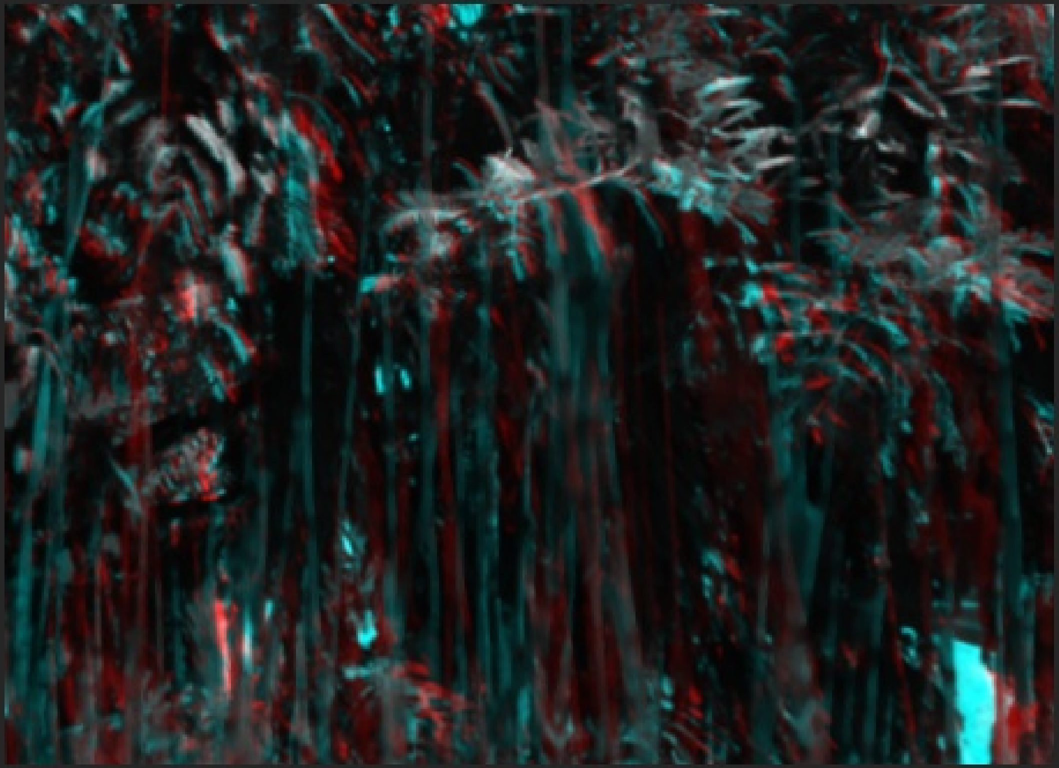


Proof of Concept: Focal stack feature



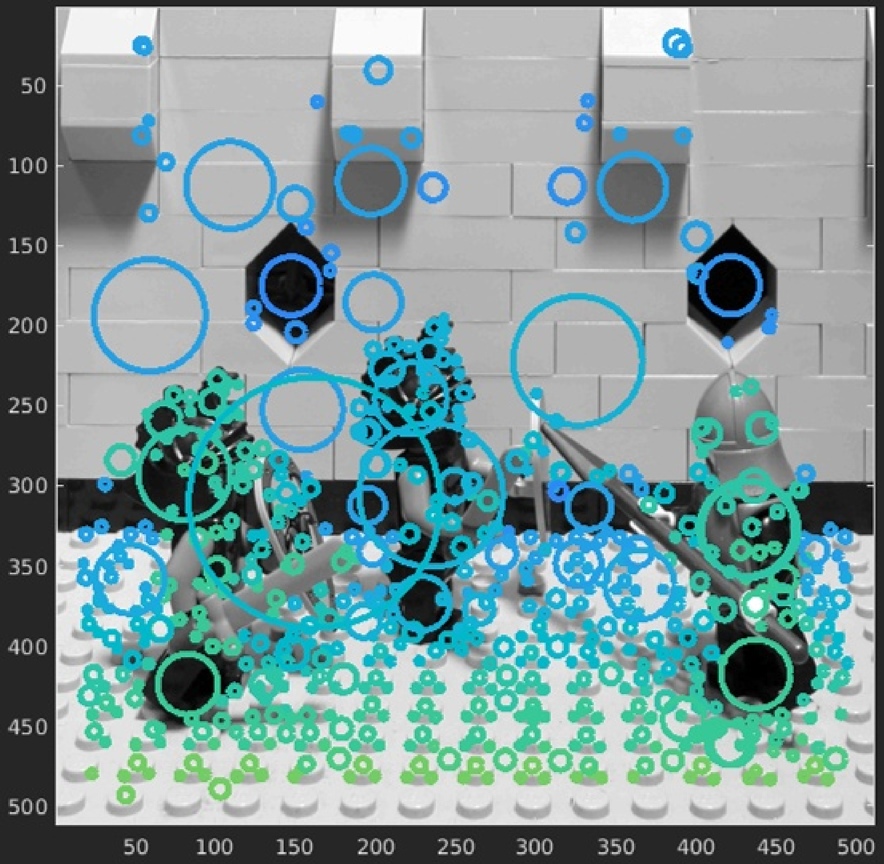
Proof of Concept: Focal stack feature

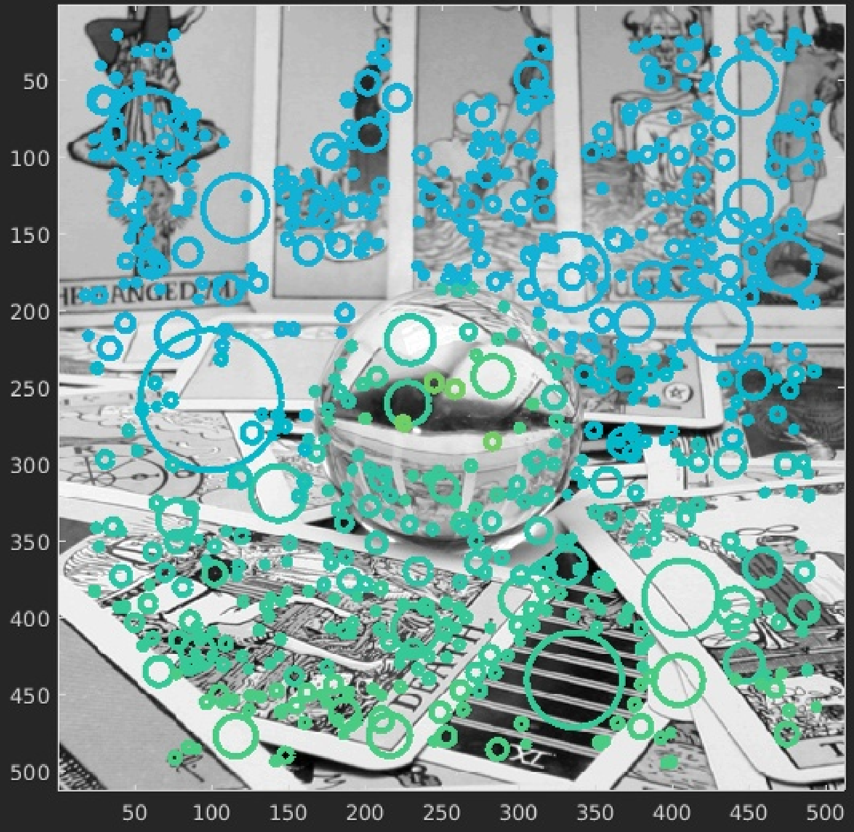
Rectified Stereo Images



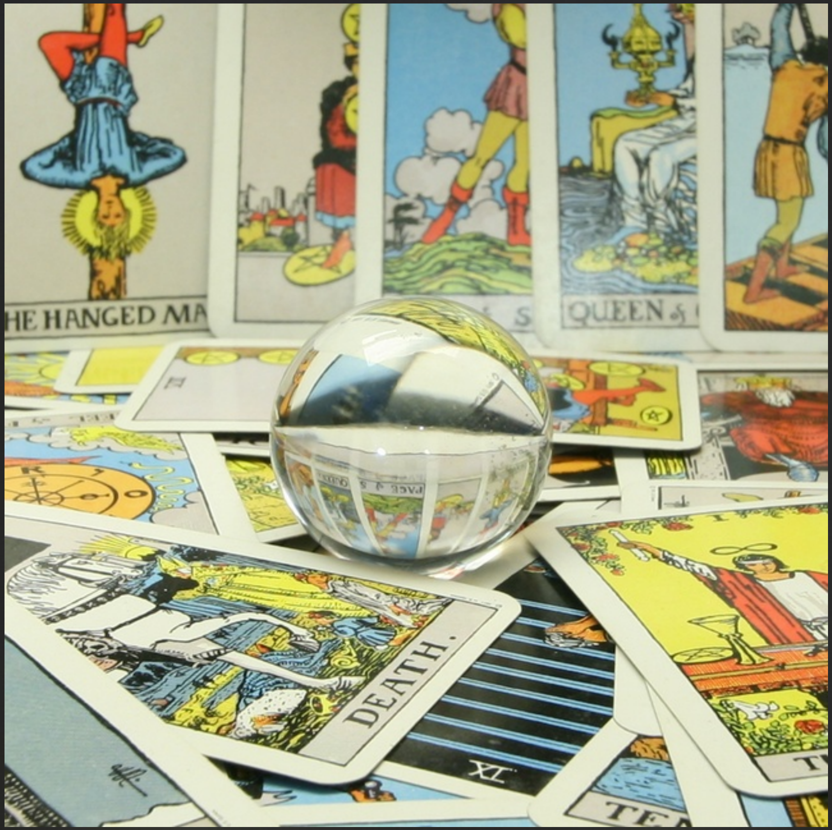
Proof-of-Concept: Slope

Depth (slope) from 2D Hough

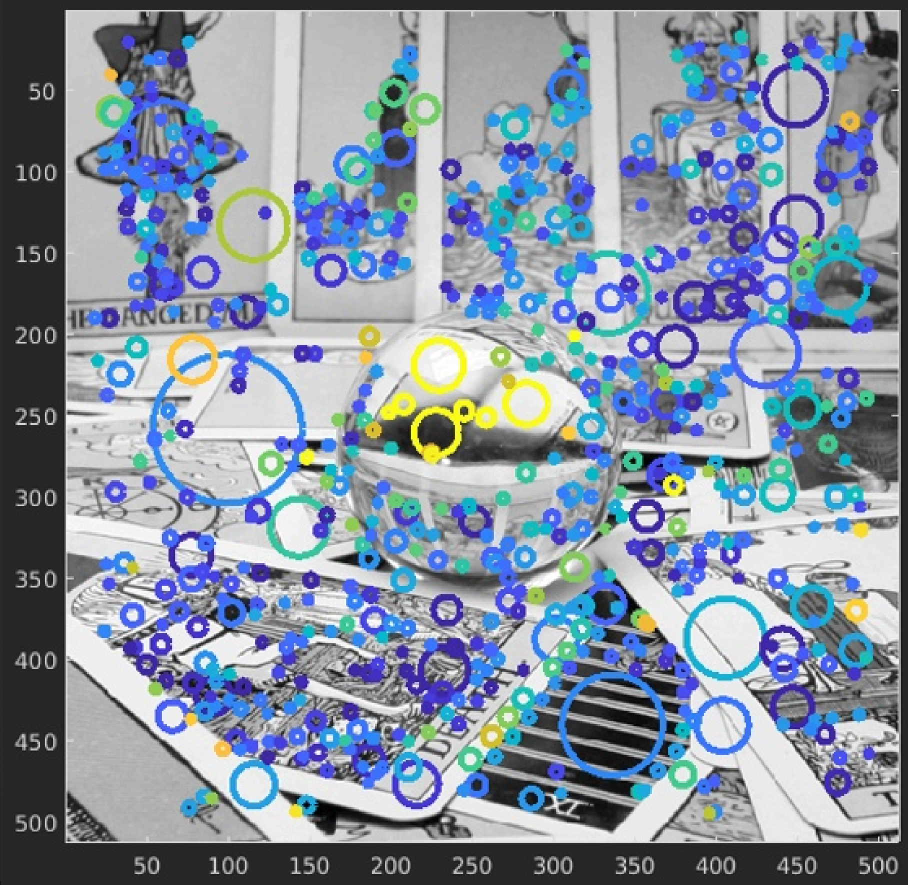




Proof-of-Concept: Refractions

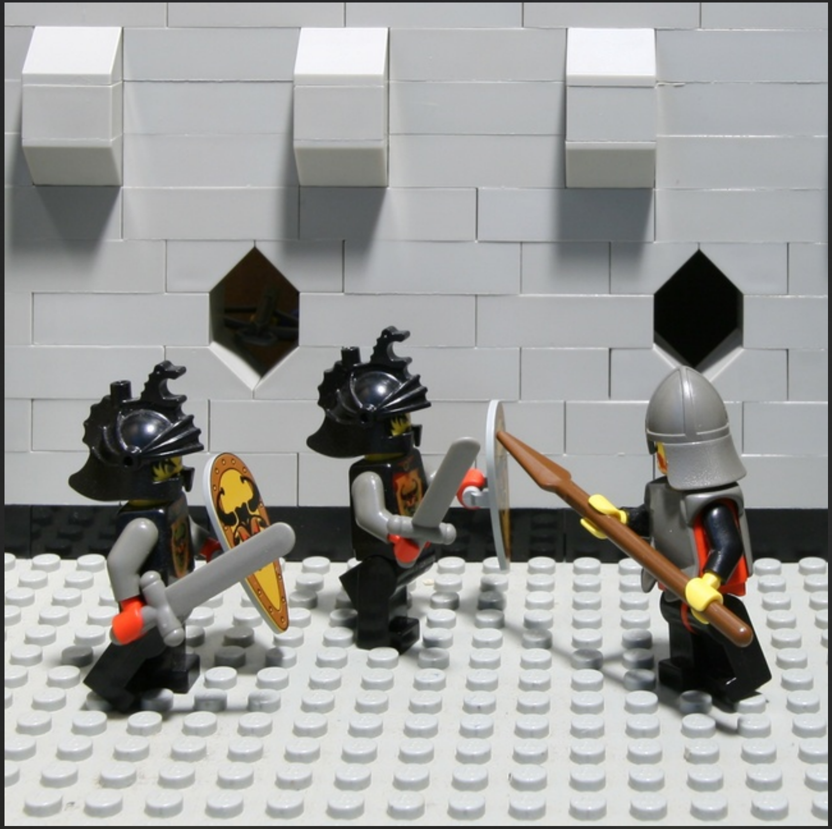


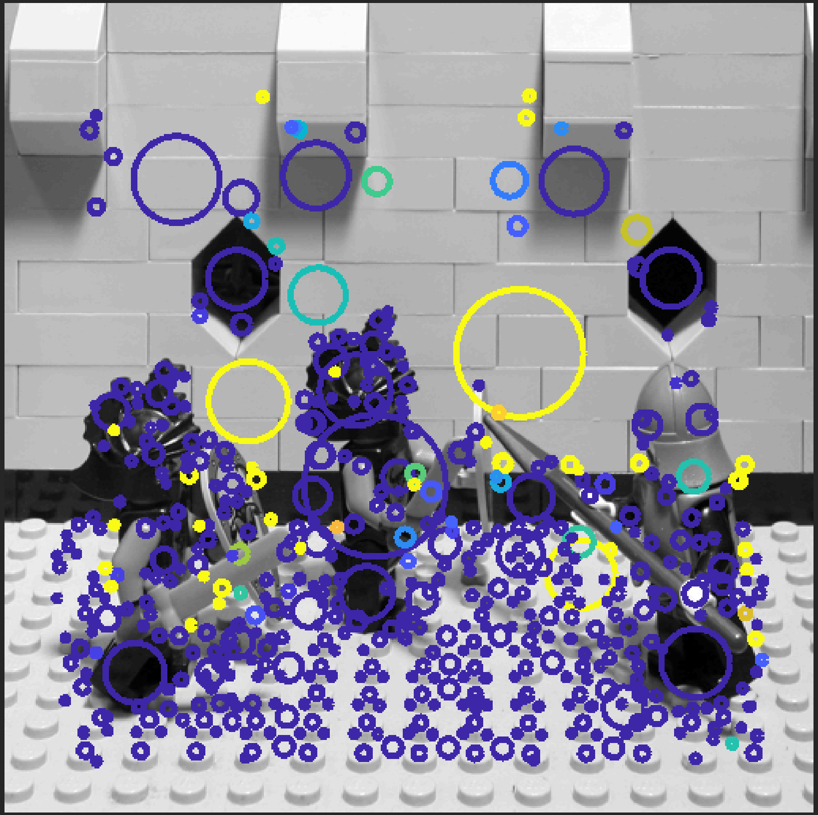
Yellow = poor feature



Proof-of-Concept: Occlusions

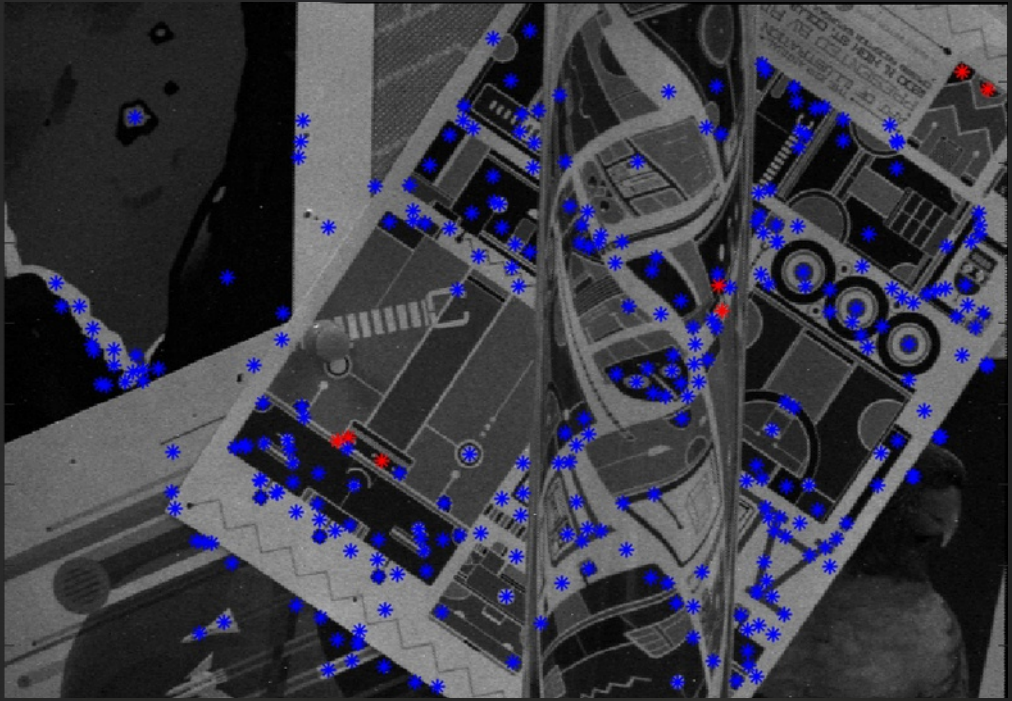
Yellow = poor feature





Proof-of-Concept: Faster Refractions

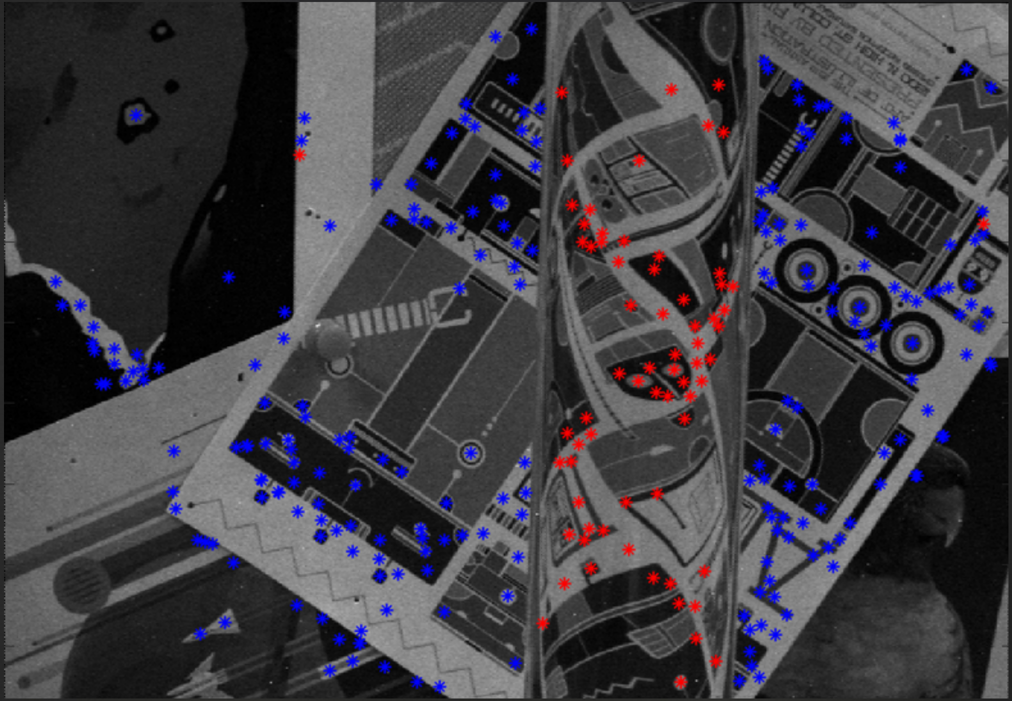
[w/Dorian Tsai, QUT, Brisbane]



Xu 2015 Transcut

Proof-of-Concept: Faster Refractions

[w/Dorian Tsai, QUT, Brisbane]

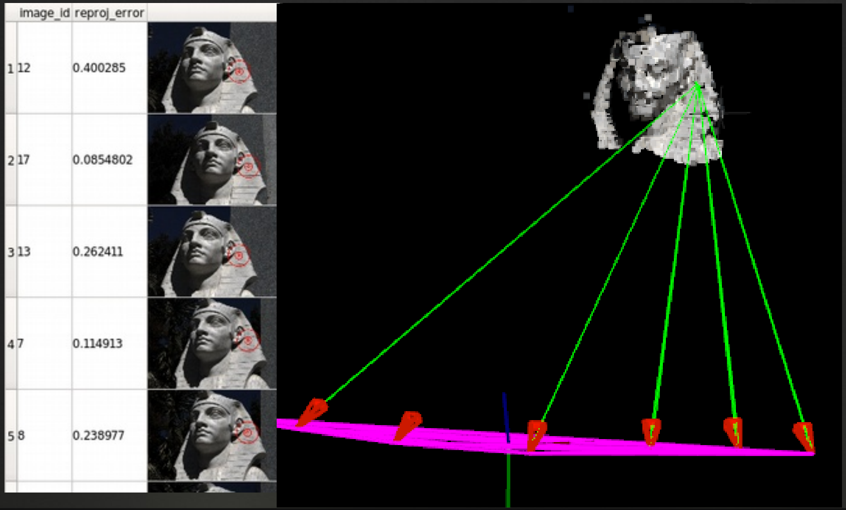


Proposed → IROS

Evaluation

Quantitative comparison to 2D SIFT and Teixeira2017

- Ground truth obtained via hand-curated SfM & Multi-View LF Dataset
- Repeatability, putative match ratio, precision, matching score, recall
- Computational complexity (FLOPS, O(.)), [speed of MATLAB implementation]



Qualitative demonstrations

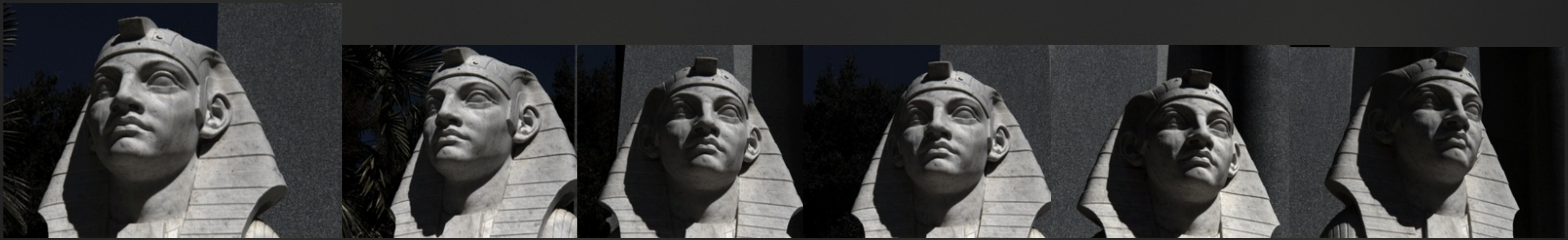
- SfM failing for 2D / naive features
- … and succeeding with LiFF features

[Teixeira2017, schonberger2017, heinly2012]

Dataset

- 4251 LFs in 31 categories
- Illum camera, varying zoom, focus, exposure
- Indoor, outdoor, easy, challenging
- 3-6 views of each scene, fixed focus/zoom
- Some revisited sites: illumination variation
- Uncalibrated camera, rough intrinsics / rectification

bamboo
batteries
benches
bikes
books
bottles
boxes
buildings
cables
cacti
chairs
coins
cups
drawers
fire_hydrants
flowers
glasses
glue
keyboards
leaves
misc
pens_and_pencils
people
phones
screws
shelf
signs
succulents
tables
tools
trees



Milestones

Slope est / feature rejecter [0.5 week]

Low-complexity 4D SIFT [1]

- Practically fast implementation [0.5]

Adaptive version [1]

Evaluation [2]

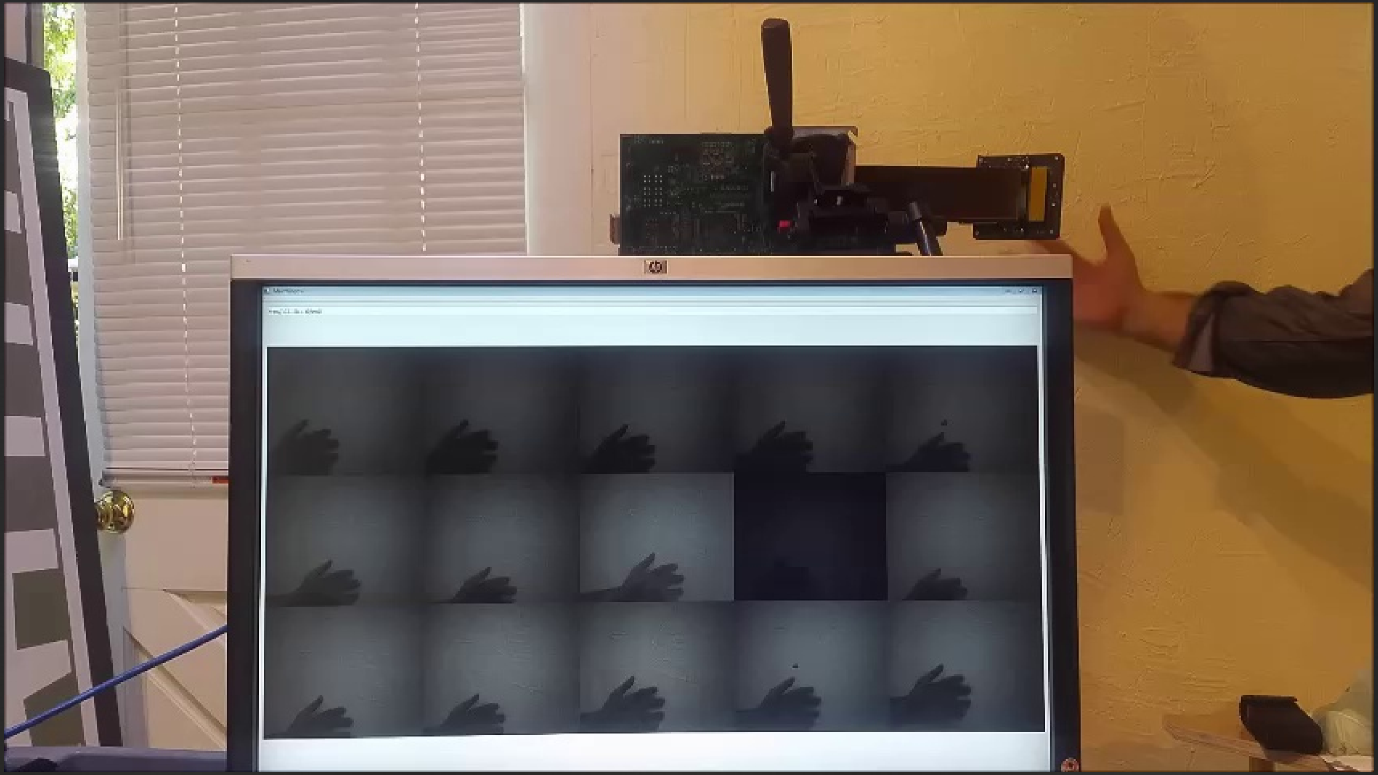
Paper [1]

Colmap ground truth [2 weeks]

Challenging example collection [1]

LF Capture

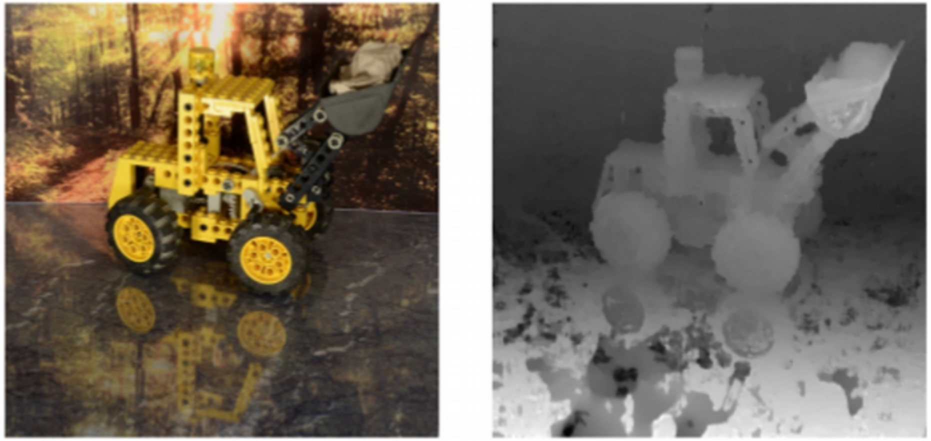
EPIImaging
Module



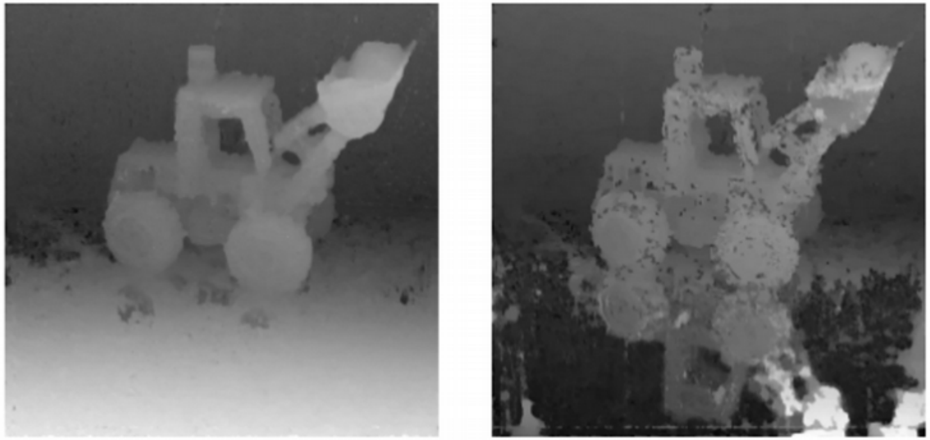
e.g. Reflection

Reflections cause spurious matches [Wanner2013]

Fix: Multi-orientation analysis



Center view and stereo reconstruction



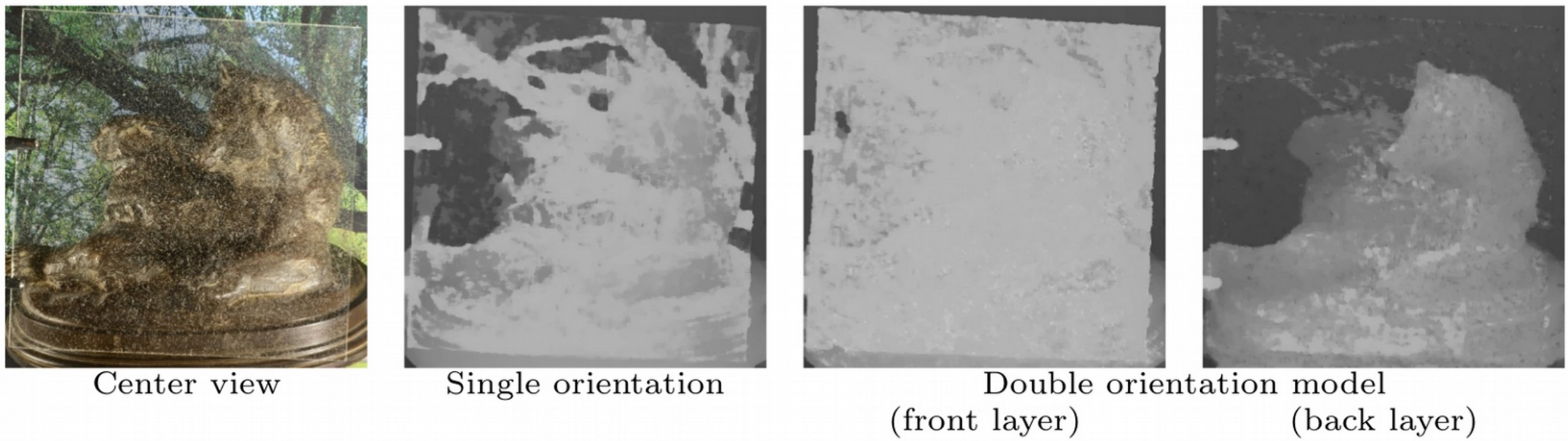
Proposed double orientation analysis

Epipolar plane image and two recovered orientations at the center location

e.g. Occlusions

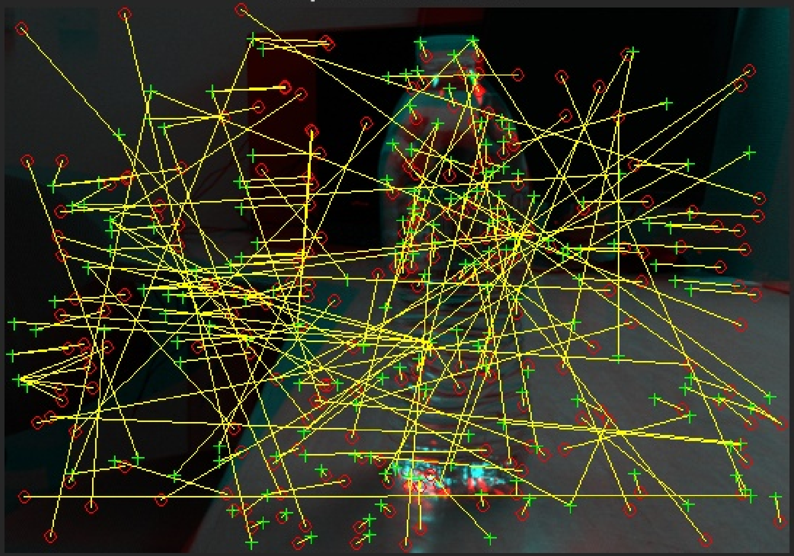
Occlusions break matching [Wanner2013]

Fix: Multi-orientation analysis

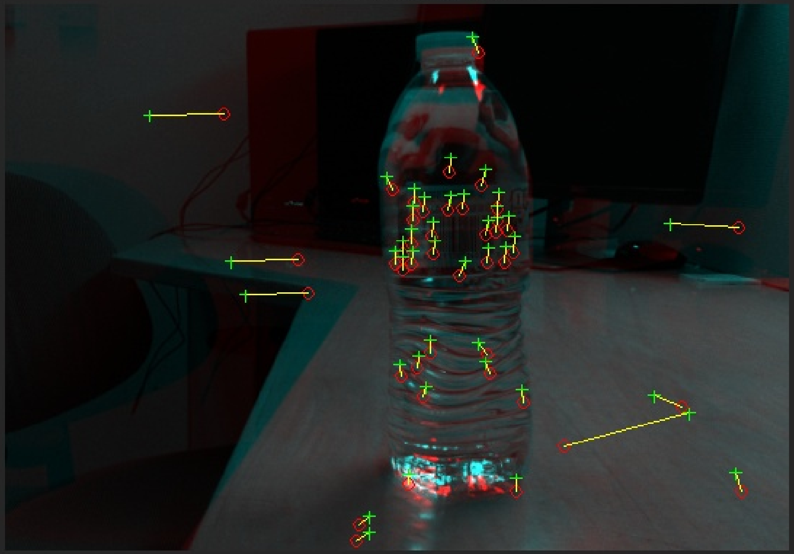


e.g. Refraction

all putative matches

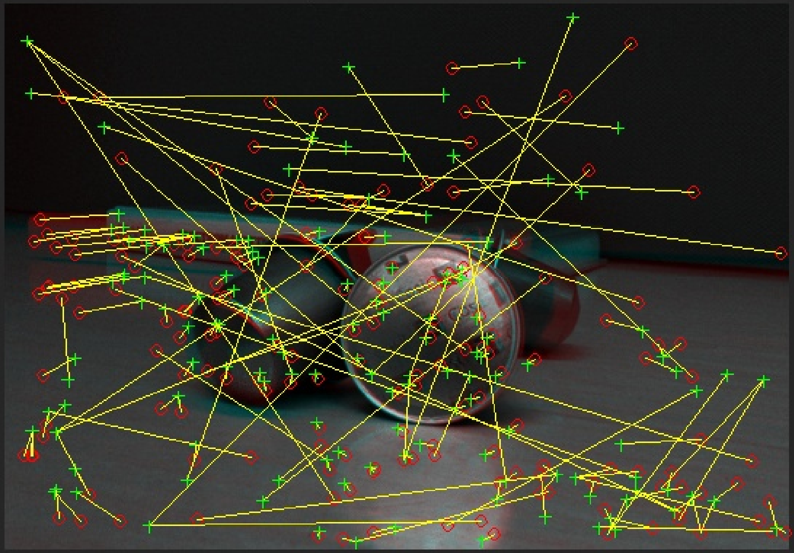


post-RANSAC matches

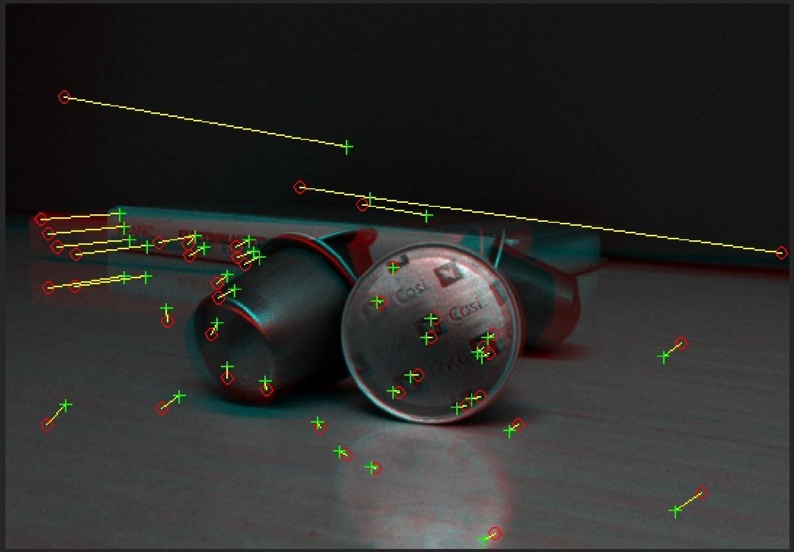


e.g. Reflection & Low Light

all putative matches

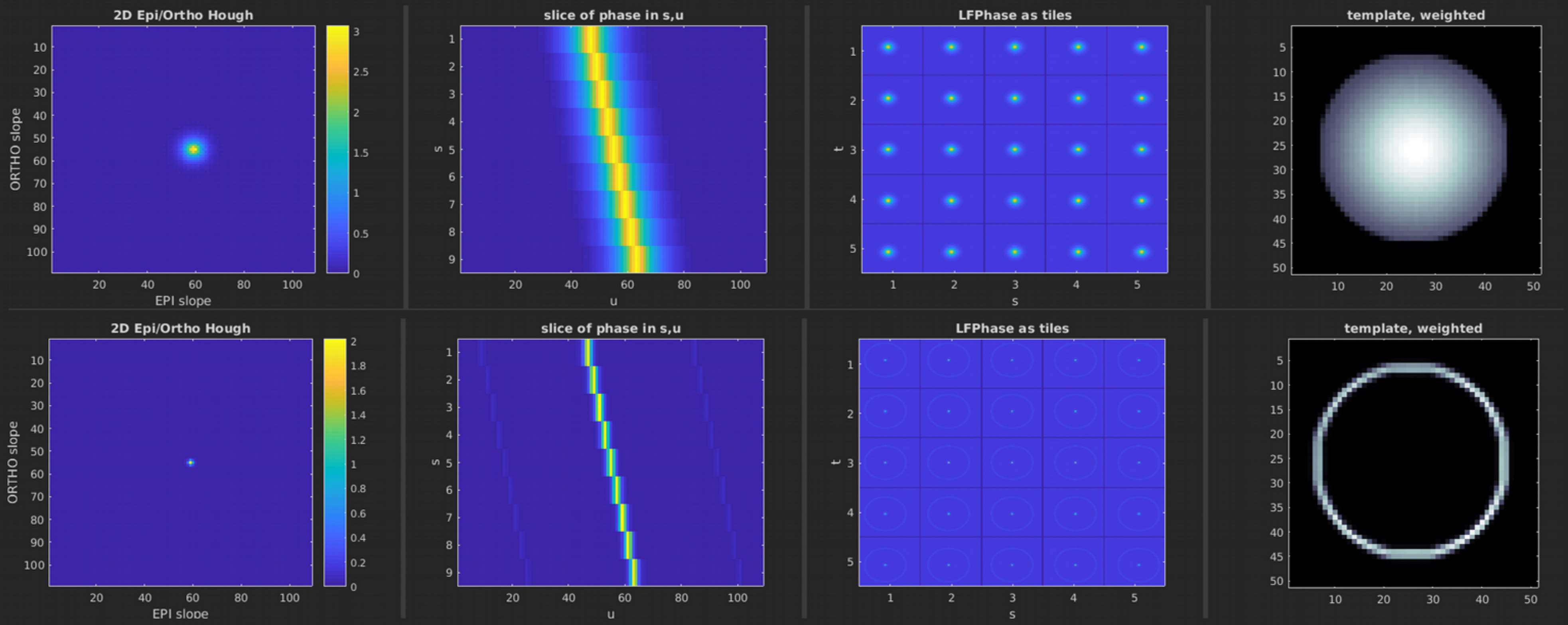


post-RANSAC matches



Refinements

Refinement1: square the voting space to increase peak contrast
Refinement2 (shown before/after here): highpass filter before cross-correlation to thin peak



Detail: Low-level Metrics

Detector Repeatability

Fraction of features correctly re-detected under camera pose change

Putative Match Ratio = #Putative Matches / #Detected Features

Fraction of detected features initially identified as a match, i.e. selectivity of matching

Precision = #Inlier Matches / #Putative Matches

How many putative matches are good

Matching Score = #Inlier Matches / #Detected Features

Number of deteced features that will result in good matches

Recall = #Inlier Matches / #True Matches

How many true matches were found

[Mikolajczyk2005, Schonberger2017, Heinly2012]

Previous Work

[Tosic 2014 “3D keypoint detection by light field scale-depth space analysis”]

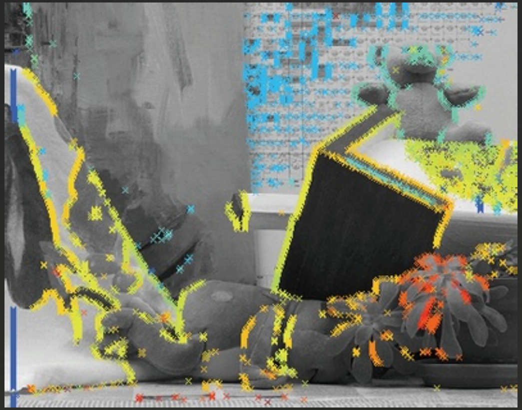
→ Detects edge keypoints, no descriptor, assumes Lambertian

[Ghasemi 2014 “Scale-invariant representation of light field images for object recognition and tracking”]

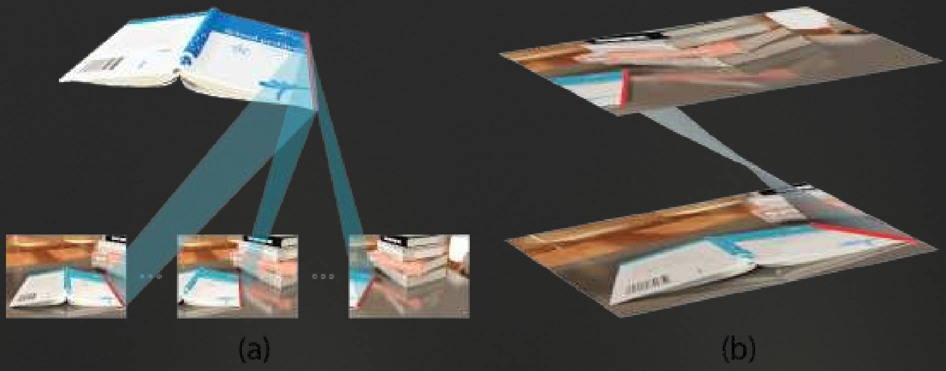
→ Global (full frame) descriptor

[Zhang 2017 “Ray Space Features for Plenoptic Structure-From-Motion”]

→ Line segment detector in all subviews



[Tosic 2014]



[Zhang 2017]

Are Hand-Crafted Features Relevant?

[Schonberger 2017 "Comparative Evaluation of Hand-Crafted and Learned Local Features"]

→ Hand-crafted modern SIFT are **faster** and **better** at reconstruction than learned features

→ Framework for evaluating features for reconstruction tasks

→ Only compares descriptors, not detectors; still a good framework for comparing

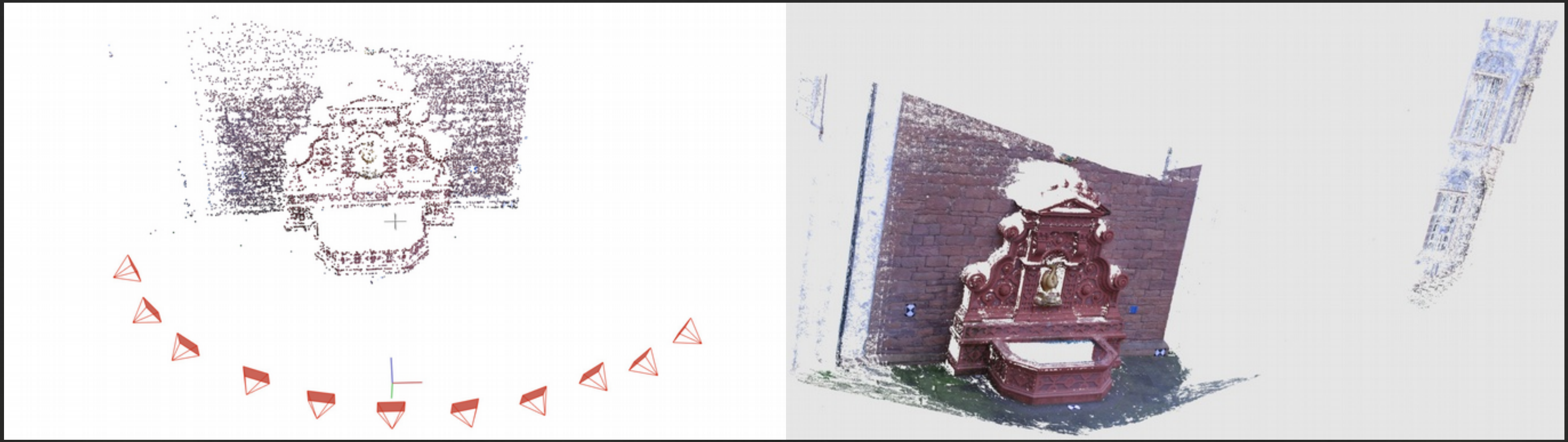


Figure 11. Sparse and dense reconstruction of Fountain for DSP-SIFT.

Is Speed Important / a Contribution?

[SIFT: Lowe 2004] “near real-time performance”, “0.3 seconds on a 2GHz Pentium 4”

IJCV: >45,000 citations

[FAST: Rosten 2006] “Machine learning for high-speed corner detection”

ECCV: >3,000 citations

[SURF: Bay 2008] “Speeded-Up Robust Features”

ECCV & CVIU: >11,000 citations

[BRISK: Leutenegger 2011] “an order of magnitude faster than SURF”

ICCV: > 2,000 citations