High-Resolution Stereo Matching

Sudipta N. Sinha

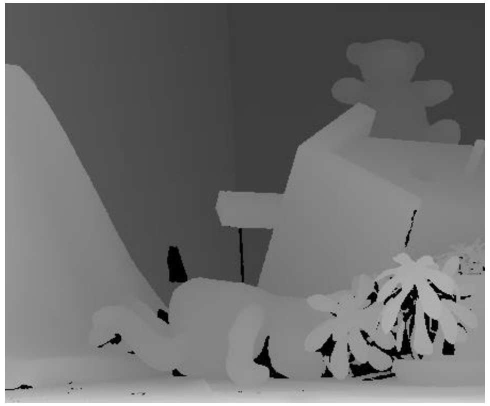
Interactive Visual Media Group
Microsoft Research

Stereo matching

- Input: rectified image pair
- Output: disparity map

SGM [Hirschmuller 2005]

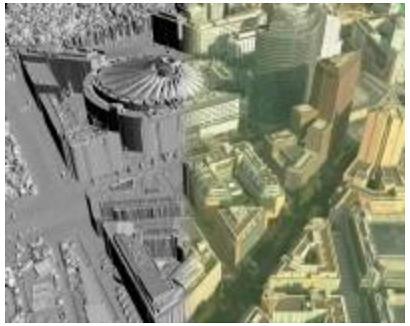
GT

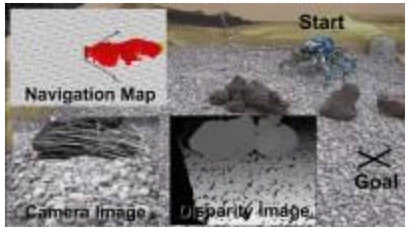


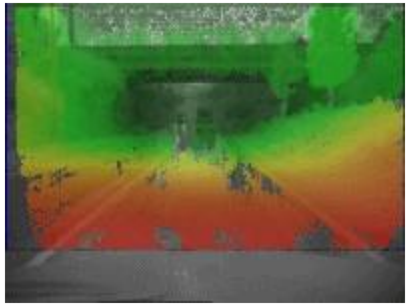
6% errors
|Δd| ≤ 1 pixel
in non-occluded
regions

Stereo – applications

- Dense 3D reconstruction
- Robot navigation
- Automated driving
- Gaming, user interfaces
- Virtual viewpoint correction
- 3D movie editing





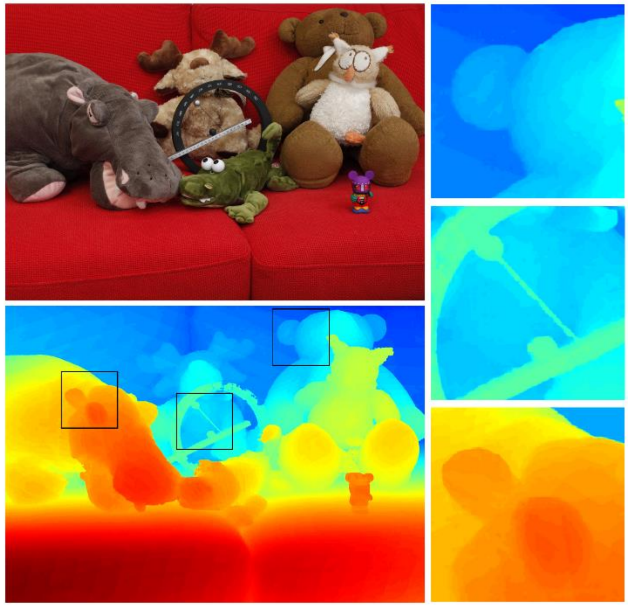


High-resolution Cameras

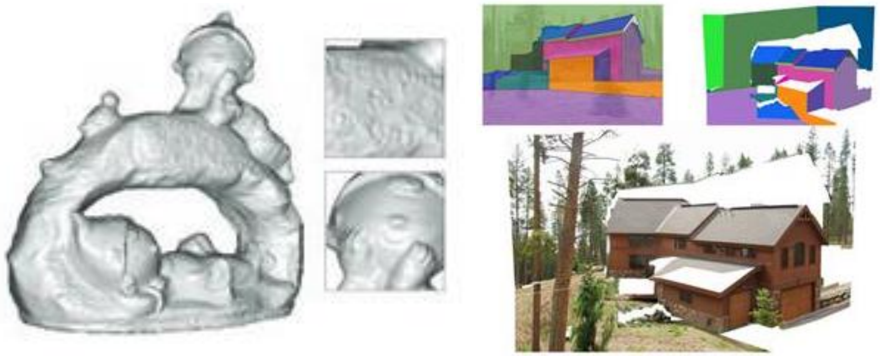
Nokia Lumia 1020
(41 MP sensor)

- High resolution cameras are everywhere
- 8+ MP on most commodity camera phones
- Impact on stereo matching techniques
  - Accuracy
  - Speed

Outline



- Stereo Matching
  - State of the art
  - High-resolution and Scalability
- High-Resolution Stereo Matching using Local Plane Sweeps
- Surface-based stereo matching
- Multi-view Stereo + Photometric Stereo



Middlebury Stereo benchmark

Images up to 450 x 375 (< 0.2 MP), D = 16 … 60

vision.middlebury.edu

stereo • mview • MRF • flow • color

Stereo | Evaluation • Datasets • Code • Submit

Middlebury Stereo Evaluation - Version 2

New features and main differences to version 1.
Submit and evaluate your own results.

☐ Open a new window for each link

Error Threshold = 1 (Error Threshold...)		Sort by nonocc				Sort by all				Sort by disc		
Algorithm	Avg.	Tsukuba ground truth			Venus ground truth			Teddy ground truth			Cones ground truth			Average percent of bad pixels (explanation)
	Rank	nonocc	all	disc	nonocc	all	disc	nonocc	all	disc	nonocc	all	disc	
ADCensus [82]	10.9	1.07 16	1.48 13	5.73 19	0.09 2	0.25 9	1.15 2	4.10 13	6.22 6	10.9 11	2.42 14	7.25 11	6.95 15	3.97
AdaptingBP [16]	14.2	1.11 19	1.37 8	5.79 21	0.10 4	0.21 6	1.44 6	4.22 15	7.06 12	11.8 15	2.48 18	7.92 23	7.32 23	4.23
CoopRegion [39]	14.8	0.87 4	**1.16** 1	4.61 4	0.11 5	0.21 5	1.54 10	5.16 26	8.31 16	13.0 21	2.79 35	7.18 10	8.01 40	4.41
RDP [87]	19.2	0.97 9	1.39 10	5.00 9	0.21 34	0.38 24	1.89 19	4.84 18	9.94 28	12.6 18	2.53 21	7.69 16	7.38 24	4.57
MultiRBF [129]	19.6	1.33 41	1.56 17	6.02 28	0.13 8	0.17 2	1.84 16	5.09 24	6.36 7	13.4 25	2.90 42	6.76 5	7.10 20	4.39
DoubleBP [34]	20.0	0.88 6	1.29 5	4.76 7	0.13 9	0.45 41	1.87 18	3.53 10	8.30 15	9.63 6	2.90 41	8.78 50	7.79 32	4.19
MDPM [140]	20.3	1.15 20	1.59 20	6.14 31	0.14 14	0.36 22	1.52 9	3.79 11	5.78 4	11.1 13	2.74 29	8.38 36	7.91 35	4.22
OutlierConf [40]	20.5	0.88 5	1.43 12	4.74 6	0.18 23	0.26 11	2.40 34	5.01 20	9.12 24	12.8 20	2.78 34	8.57 41	6.99 16	4.60
AdaptiveGF [127]	24.1	1.04 12	1.53 14	5.62 14	0.17 22	0.41 32	1.98 22	5.71 35	11.3 41	14.3 32	2.44 16	8.22 30	7.05 19	4.98

KITTI benchmark

Image size 1241 x 376 (< 0.5 MP), D ≈ 70…150

The KITTI Vision Benchmark Suite
A project of Karlsruhe Institute of Technology and Toyota Technological Institute at Chicago

Toyota Technological Institute Chicago | KIT Karlsruhe Institute of Technology

home | setup | stereo | flow | odometry | object | tracking | road | raw data | submit results | jobs

Andreas Geiger (MPI Tübingen) | Philip Lenz (KIT) | Christoph Stiller (KIT) | Raquel Urtasun (University of Toronto)

Stereo Evaluation

Rank	Method	Setting	Code	Out-Noc	Out-All	Avg-Noc	Avg-All	Density	Runtime	Environment	Compare
1	SceneFlow			**2.98 %**	**3.97 %**	**0.8 px**	1.0 px	100.00 %	35 s	1 core @ 3.5 Ghz (C/C++)	
Anonymous submission											
2	PCBP-SS			3.40 %	4.72 %	0.8 px	1.0 px	100.00 %	5 min	4 cores @ 2.5 Ghz (Matlab + C/C++)	
K. Yamaguchi, D. McAllester and R. Urtasun: Robust Monocular Epipolar Flow Estimation. CVPR 2013.											
3	gtRF-SS			3.83 %	4.59 %	0.9 px	1.0 px	100.00 %	1 min	1 core @ 2.5 Ghz (Matlab + C/C++)	
Anonymous submission											
4	StereoSLIC			3.92 %	5.11 %	0.9 px	1.0 px	99.89 %	2.3 s	1 core @ 3.0 Ghz (C/C++)	
K. Yamaguchi, D. McAllester and R. Urtasun: Robust Monocular Epipolar Flow Estimation. CVPR 2013.											
5	PR-Sf+E			4.02 %	4.87 %	0.9 px	1.0 px	100.00 %	200 s	4 cores @ 3.0 Ghz (Matlab + C/C++)	
C. Vogel, S. Roth and K. Schindler: Piecewise Rigid Scene Flow. International Conference on Computer Vision (ICCV) 2013.											
6	PCBP			4.04 %	5.37 %	0.9 px	1.1 px	100.00 %	5 min	4 cores @ 2.5 Ghz (Matlab + C/C++)	
K. Yamaguchi, T. Hazan, D. McAllester and R. Urtasun: Continuous Markov Random Fields for Robust Stereo Estimation. ECCV 2012.											

Limitations

- 100+ new algorithms published (benchmarked) since 2002.
  - Middlebury: focus on accuracy
  - KITTI: focus on robust performance

Neither require high accuracy on hi-res images

- Most of the existing methods do not scale.

Limitations

- Searching full disparity space requires
  O(P*D) time
  = O(s^3) for image size s

Middlebury Teddy	KITTI
0.16MP x 60 = 10 Mdisp.	0.5MP x 80 = 40 Mdisp.

Middlebury New	Disney Mansion
6MP x 256 = **1.5 Gdisp.**	19MP x 1000 = **19 Gdisp.**

New datasets and benchmarks

- Need more challenging datasets for algorithm design
- Middlebury Stereo Eval v.3 is underway *
- 30 new datasets

  - a subset is discussed here -- “Middlebury New 7”

* The benchmark is not public yet.

  If interested, contact Daniel Scharstein (schar@middlebury.edu)


Middlebury
Teddy
Middlebury
New
KITTI
Disney

Scalability

- (Most) existing methods are O(P*D), or O(P*D^2)
- These methods do not scale
- Ideally, want O(P)

P: pixels
D: disparities

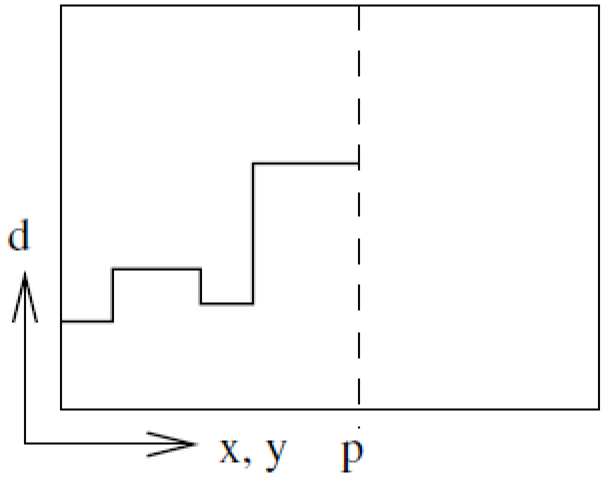
- Key Issues:
  - Does higher resolution even help?
  - Does it make sense to enumerate disparities ?
  - Coarse to fine strategies ?
  - Practical optimization techniques

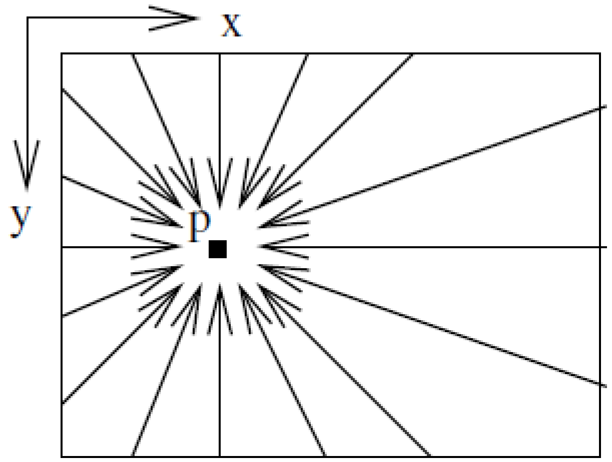
Promising Approaches

- Efficient approximate energy minimization
  - Only run at low-res; Upsample and refine disparities
    [Ferst et. al. 2013, Ma et. al. 2013]
  - Semi-global Matching (SGM)
    [Hirschmüller 2005]
- Avoid exploring the whole DSI
  - Coarse-to-fine [long tradition]
  - Seed & Grow [Cech & Sara 2007, …]
  - PatchMatch stereo [Bleyer et al. 2011]
  - ELAS [Geiger et al. 2010]
  - Local Plane Sweep Stereo [this talk]

Semi-global Matching

[Hirschmüller 2005]





- Aggregate along 1D minimum cost paths ending at pixel p
- Only cost of this path needed; not the path itself
- Efficiently computed via message-passing.
- Winner-take-all disparity selection

Efficient High-Resolution Stereo Matching using Local Plane Sweeps

Will be presented at CVPR 2014

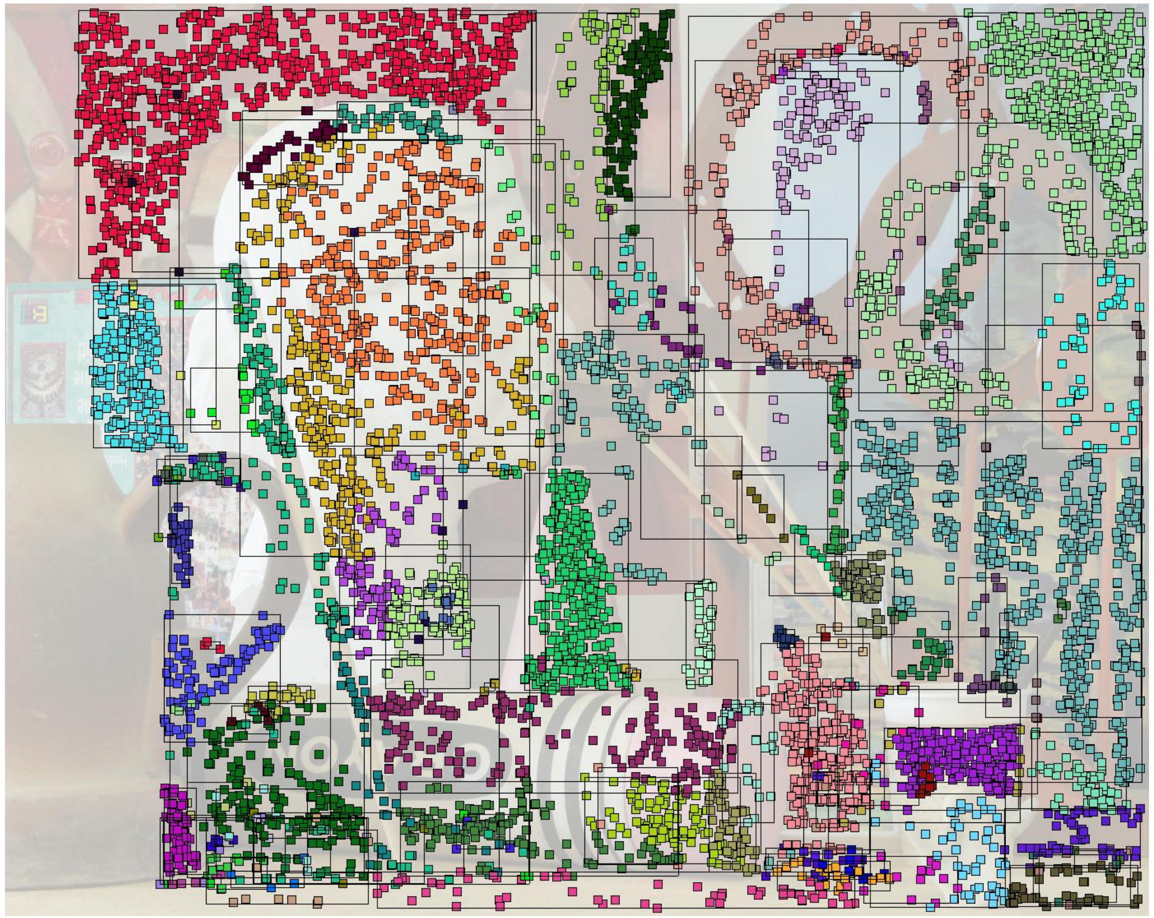
Daniel Scharstein

Rick Szeliski

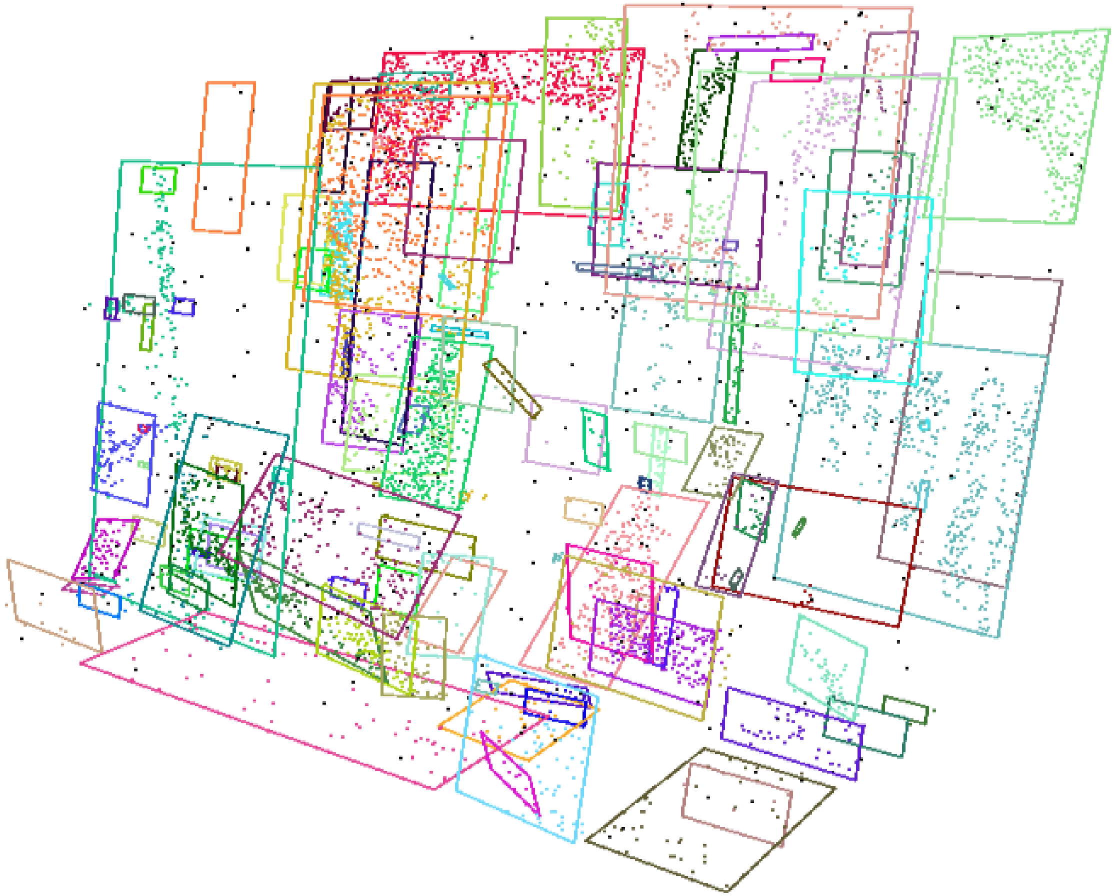
Local Plane Sweep Stereo

- Sparse feature matching; refine vertical disparities
- Generate plane hypotheses (with unknown extents)

Plane hypothesis generation



Plane Labels



Local Plane Sweep Stereo

- Sparse feature matching; refine vertical disparities
- Generate plane hypotheses (with unknown extents)
- Perform local plane sweeps (LPS) around hypothesized planes

  - local stereo problem with narrow disparity range; solved using SGM

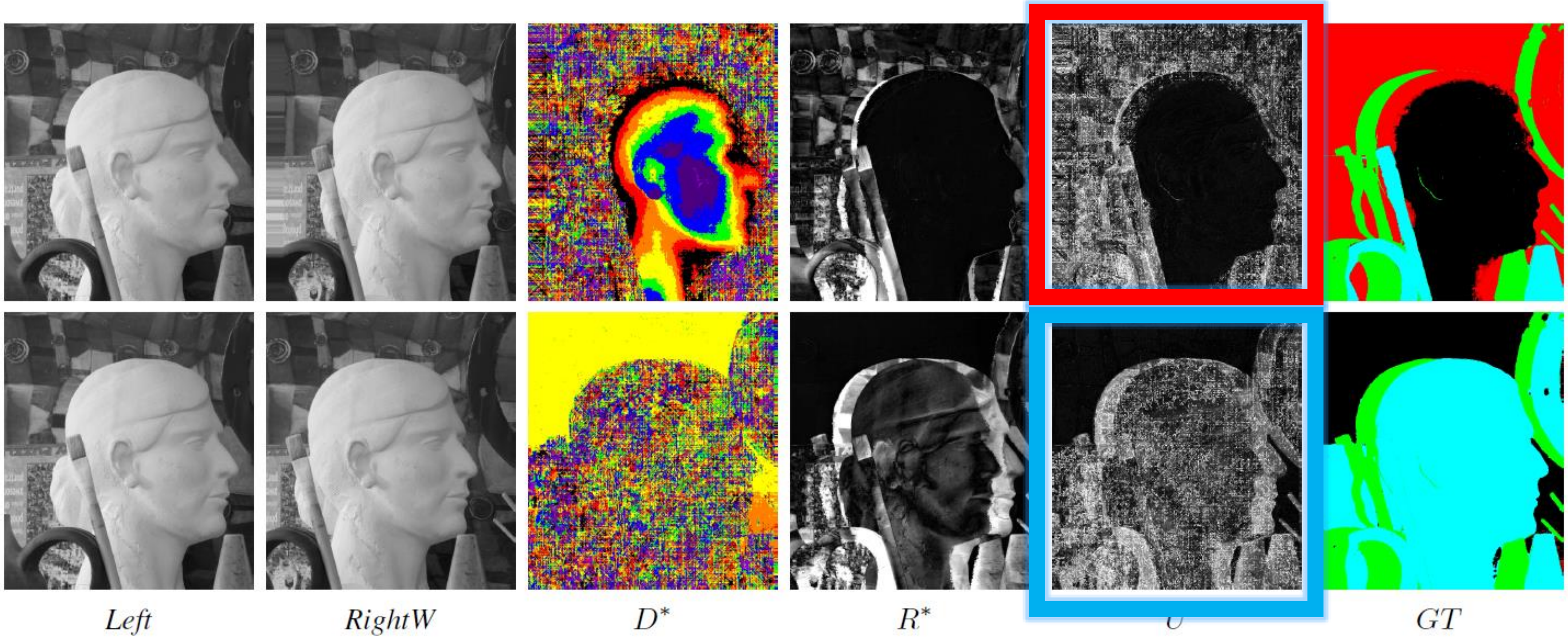
Local Plane Sweeps

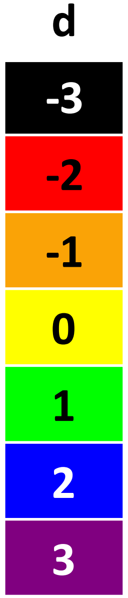
Cost Map

In-range disparities (ground truth)

Plane 2

Local Plane Sweep solution





Plane 1

Local Plane Sweep solution

Local Plane Sweep Stereo

- Sparse feature matching; refine vertical disparities
- Generate plane hypotheses (with unknown extents)
- Perform local plane sweeps (LPS) around hypothesized planes

  - local stereo problem with narrow disparity range; solved using SGM
- Tile structure

  - Perform LPS on tiles and propagate planes to adjoining tiles

Proposal Propagation

- Initial proposals:

  - planes π with feature points per tile

- Repeat nR=3 times:

  - Compute surfaces via local plane sweeps per tile
  - Update winner-take-all label map (L_{WTA})
  - Use L_{WTA} to predict well-supported plane proposals
  - Propagate these planes to the neighboring tiles

Local Plane Sweep Stereo

- Sparse feature matching; refine vertical disparities
- Generate plane hypotheses (with unknown extents)
- Perform local plane sweeps (LPS) around hypothesized planes

  - local stereo problem with narrow disparity range; solved using SGM
- Tile structure

  - Perform LPS on tiles and propagate planes to adjoining tiles
- Global optimization

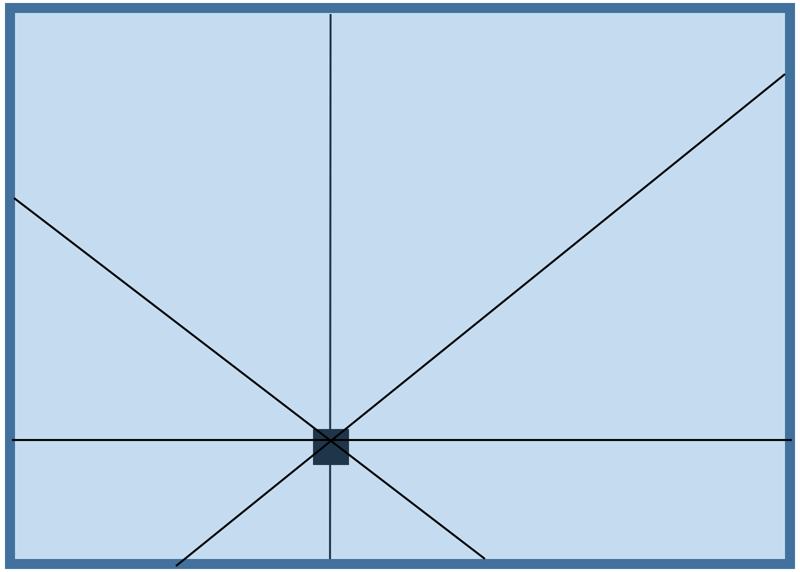
  - Assign pixels to surface proposals

  - Fast approximate energy minimization (via SGM)

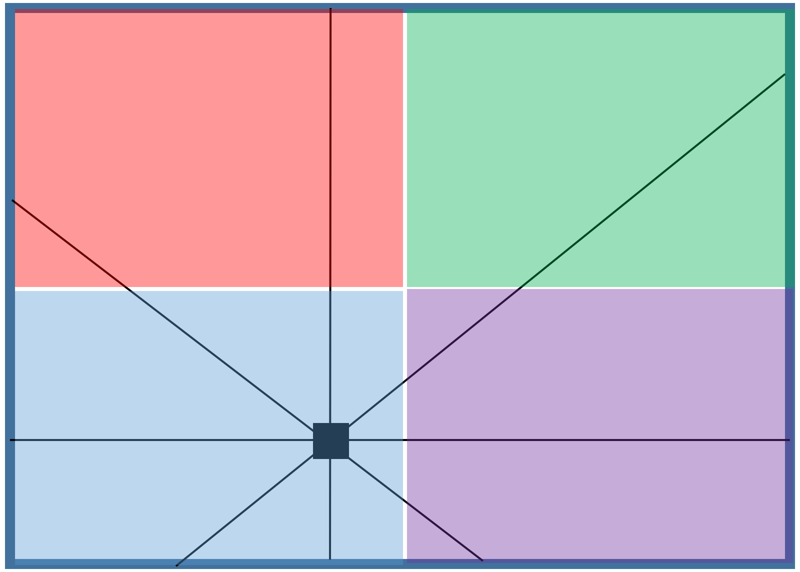
  - Extend SGM to exploit tile structure and sparse label sets

Global Optimization (via SGM)

- Message passing on 1D paths (8 directions)



SGM
- fixed label set at all pixels



LPS
- Label sets vary across tiles

Disparity Selection

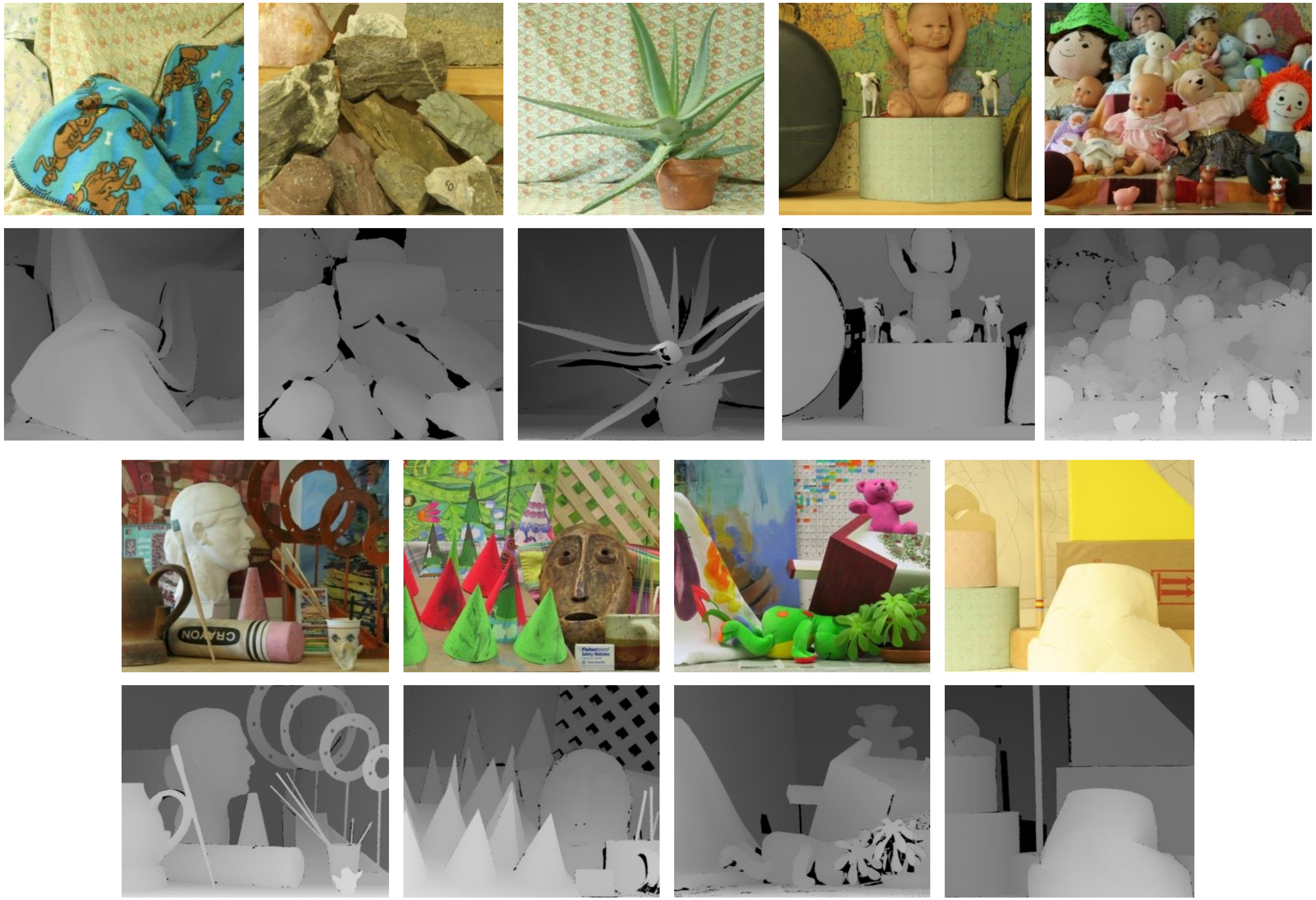
- Original SGM:
  - Aggregated costs
  - WTA at every pixel

- LPS:
  - Aggregated costs
  - Top-m candidates at every pixel (m = 2)
  - Median filter on candidates within a small window.

Experiments

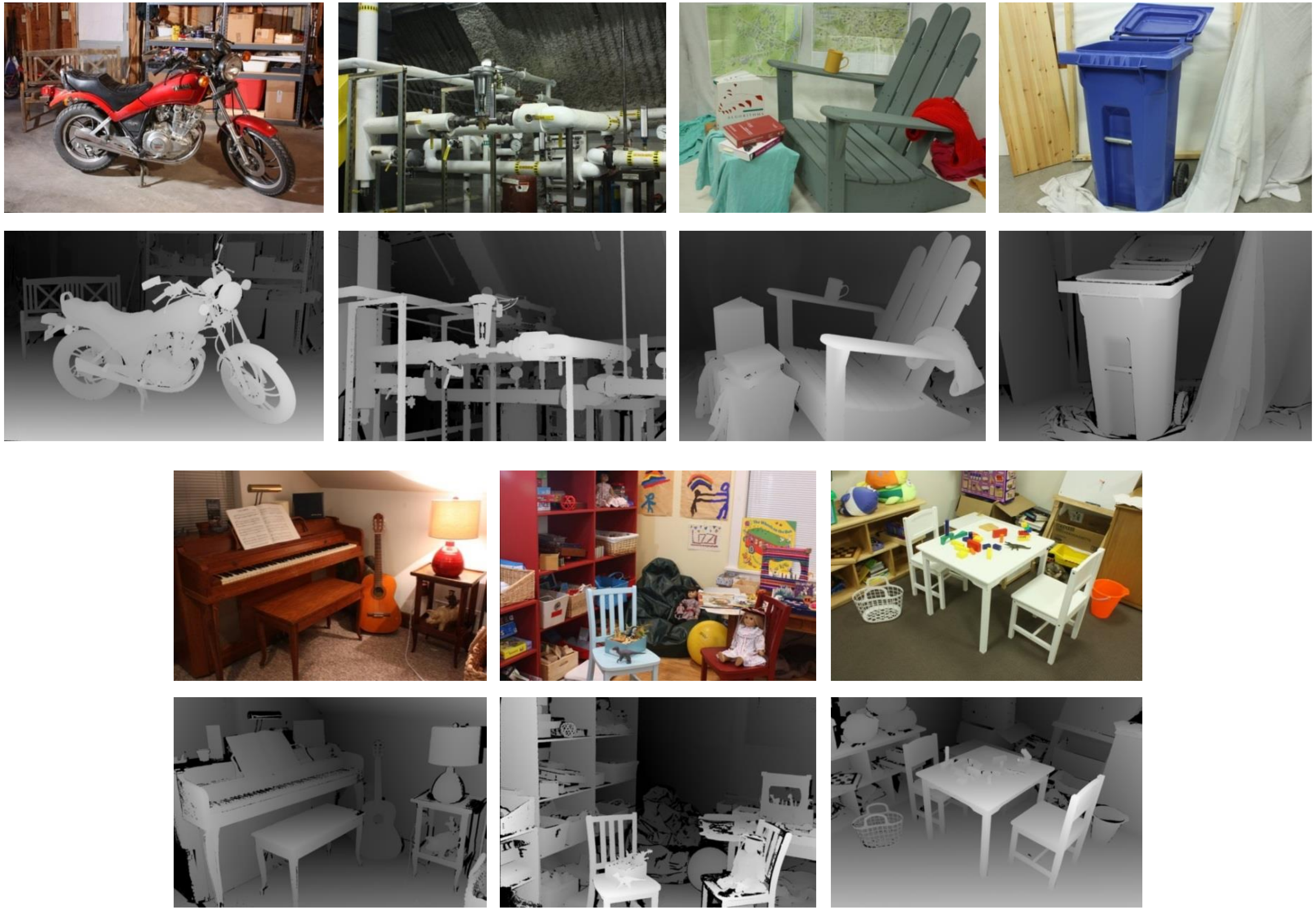
- Evaluation:
  - PatchMatch Stereo [Bleyer et al. 2011]
  - SGM (our impl.)
  - SGM-HH [Hirschmüller 2005]
  - ELAS [Geiger et al. 2010]
  - LPS
- Metric:
  - 1 and 2 pixel disparity error at non-occluded pixels.

Midd9 (1.4 – 2.7 MP)



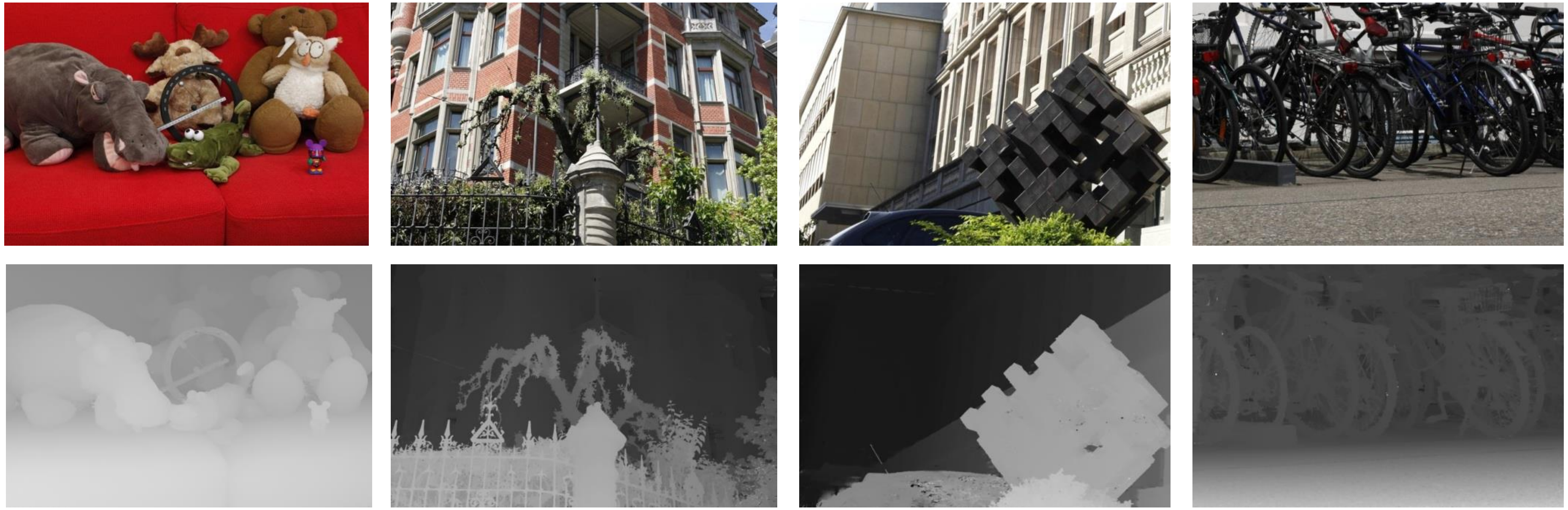
Subset of full-resolution 2003-2006 Middlebury datasets

MiddNew7 (5.1 – 6.0 MP)



Subset of new 2011-2014 Middlebury training sets

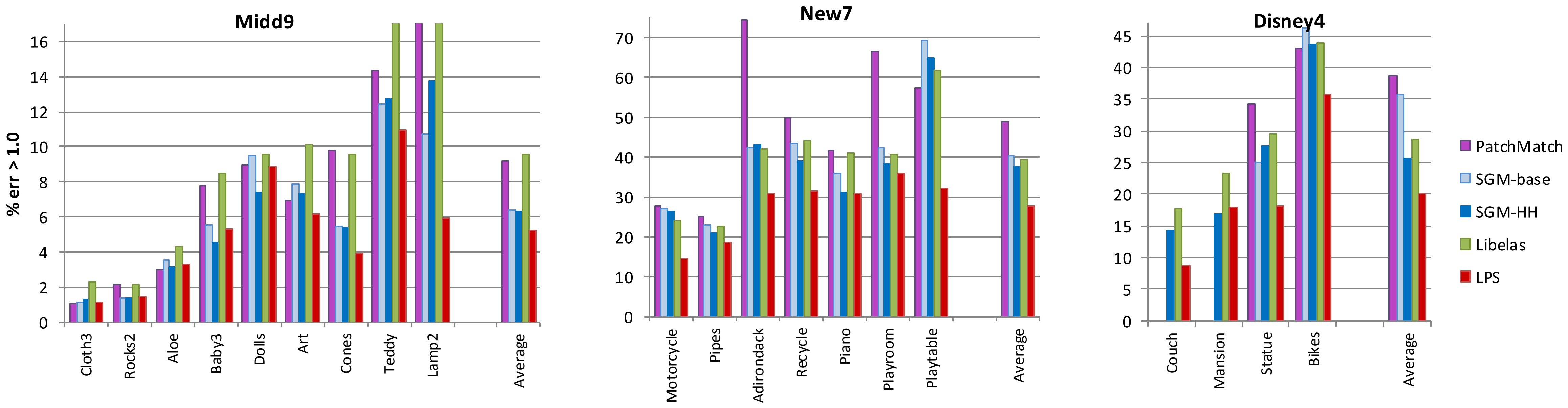
Disney4 (4.5 – 19 MP)



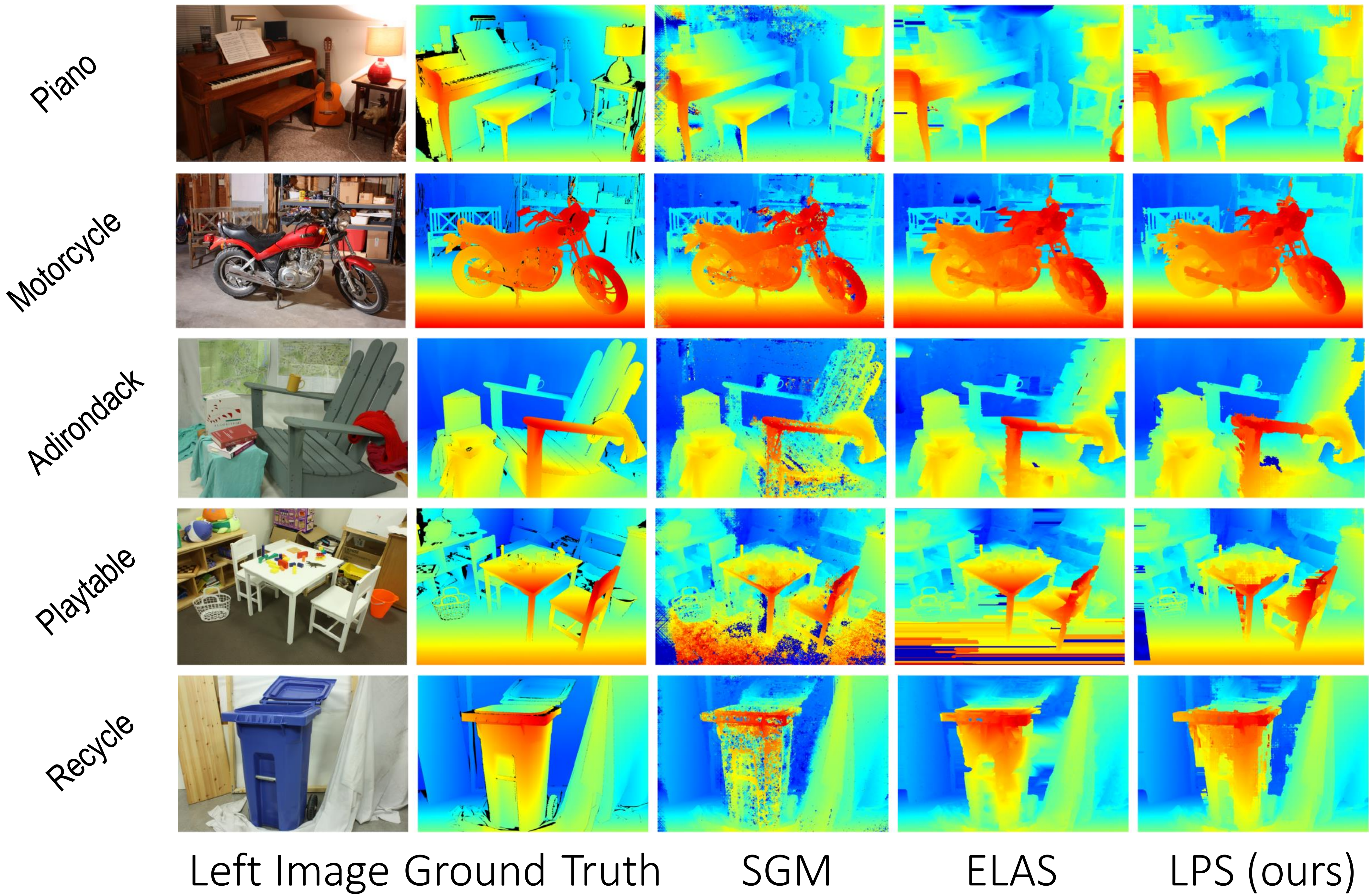
C. Kim, H. Zimmer, Y. Pritch, A. Sorkine-Hornung, and M. Gross
Scene reconstruction from high spatio-angular resolution light fields
SIGGRAPH 2013

* We treat their results (computed from 100 images) as GT

Results – accuracy



Bad pixels (%), thresh = 1.0 pixel

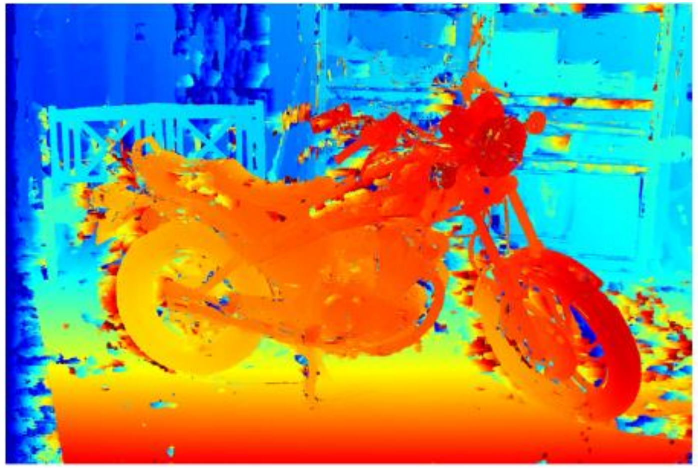

Piano
Motorcycle
Adirondack
Playtable
Recycle
Left Image
Ground Truth
SGM
ELAS
LPS (ours)

Error maps (Motorcycle)

Ground Truth

Occlusion Mask

PatchMatch



3330 seconds

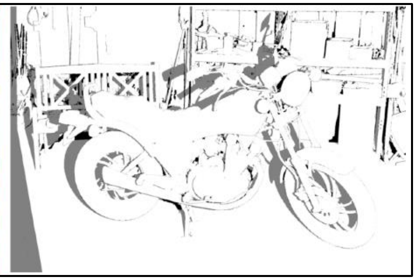
err1 = 33.8 %

err2 = 24.2 %

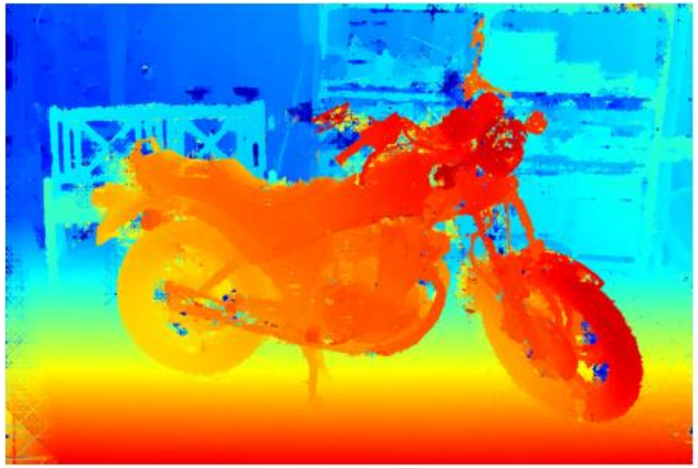
Error maps (Motorcycle)

Ground Truth

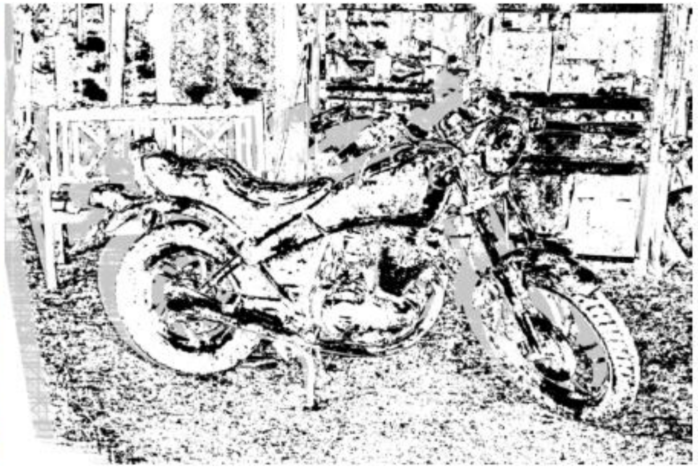
Occlusion Mask



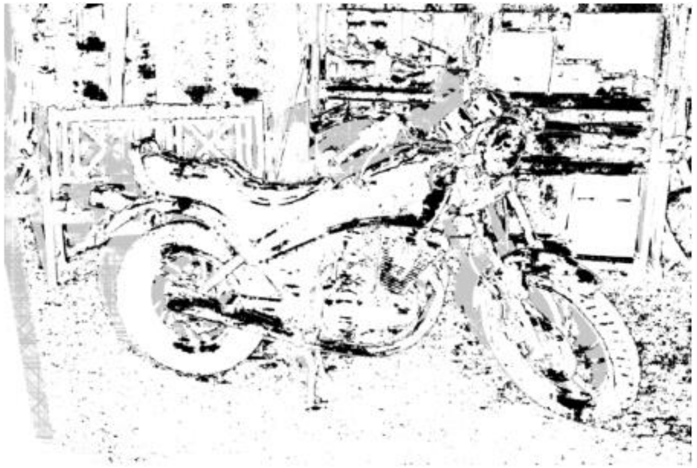
SGM



51.4 seconds



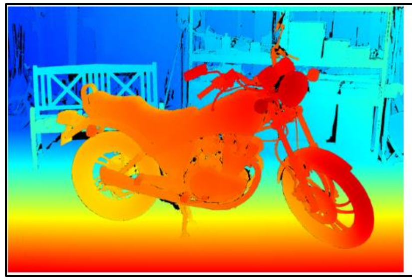
err1 = 29.3 %



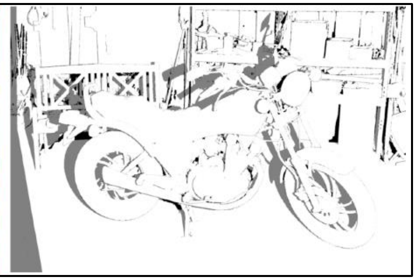
err2 = 15.1 %

Error maps (Motorcycle)

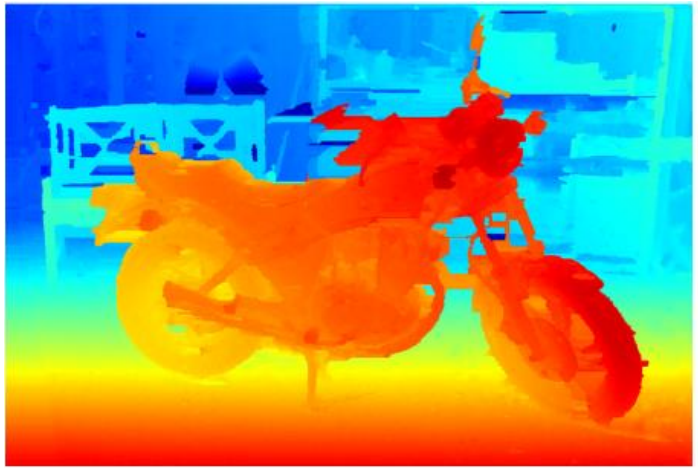
Ground Truth



Occlusion Mask

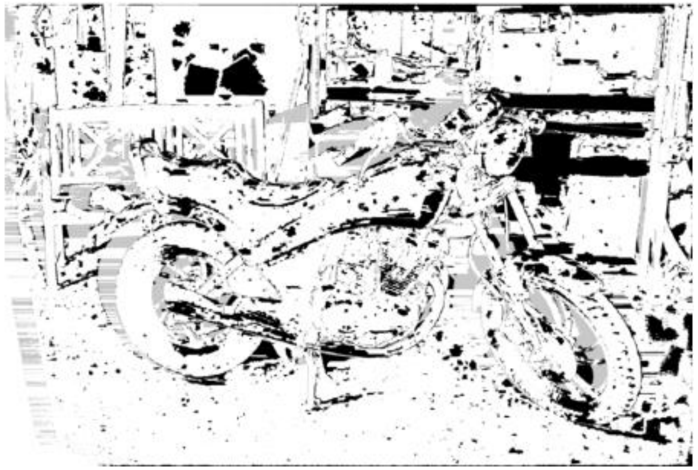


ELAS



5.0 seconds

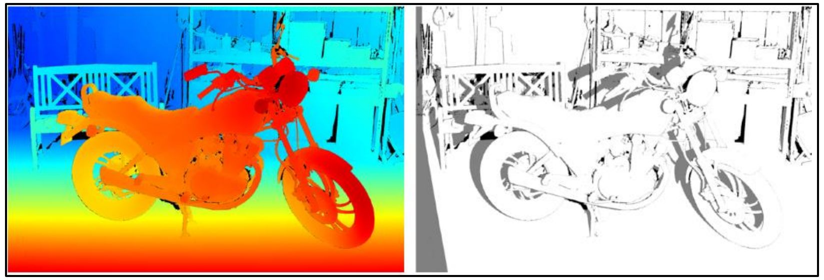
err1 = 34.0 %



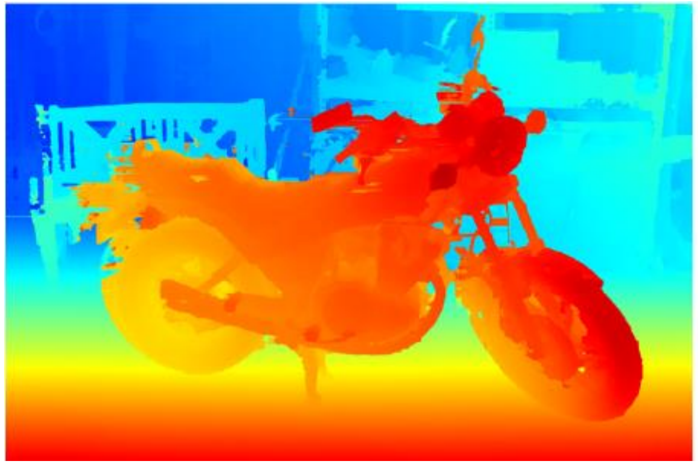
err2 = 19.1 %

Error maps (Motorcycle)

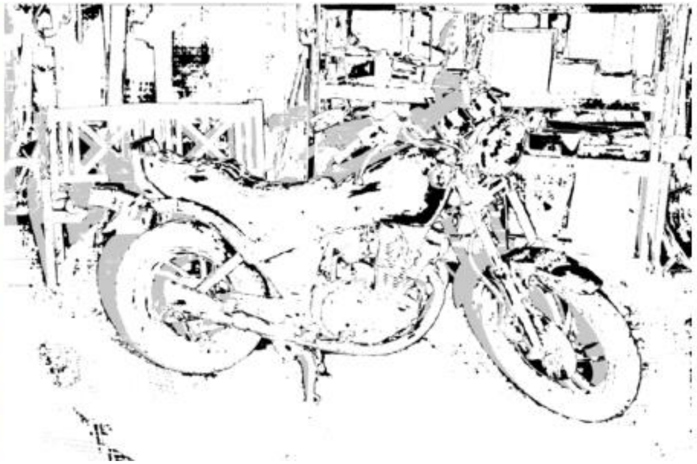
Ground Truth Occlusion Mask



LPS



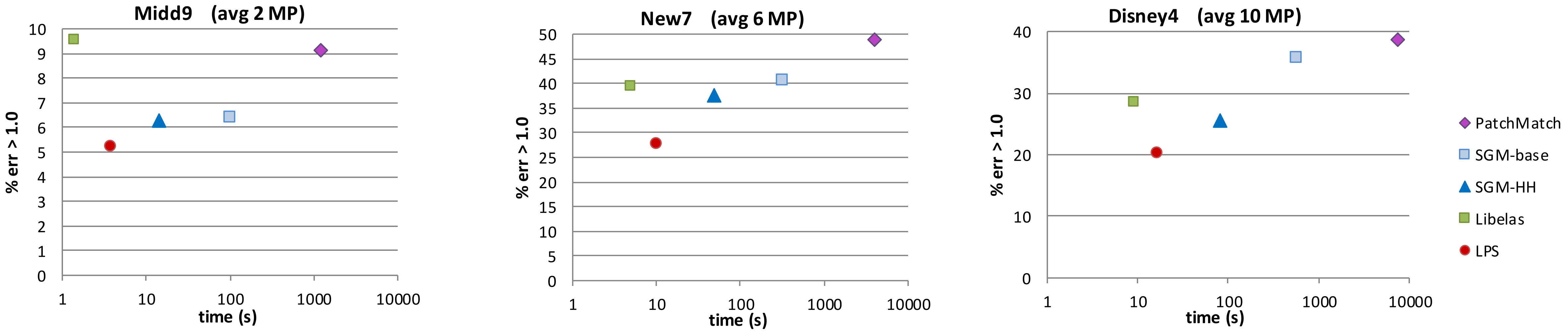
9.6 seconds



err1 = 12.2 %

err2 = 6.5 %

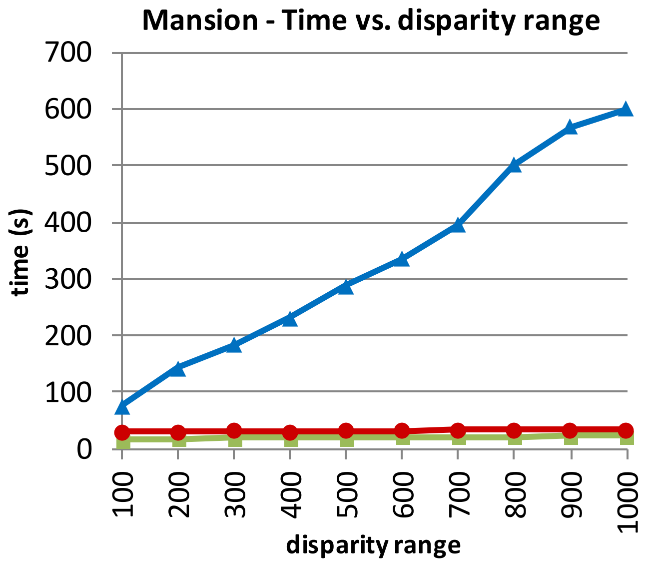
Results – Accuracy vs. Runtime



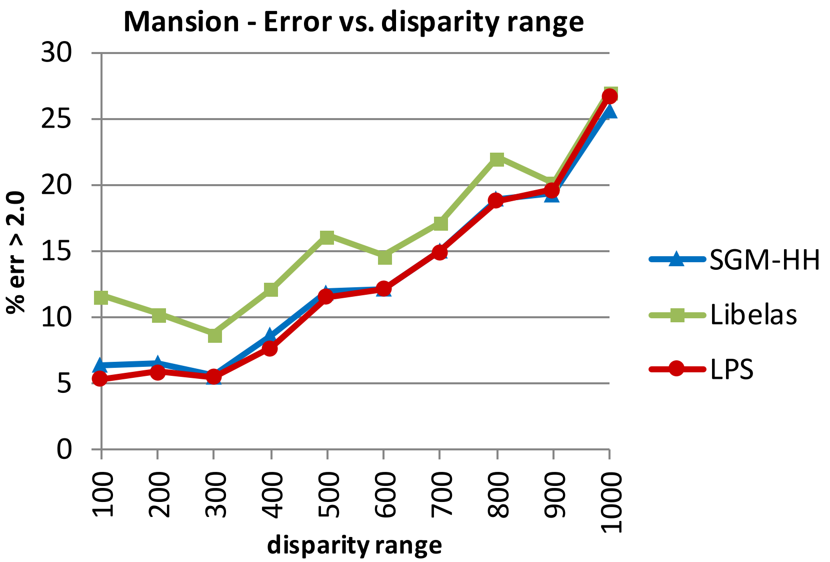
Avg. error vs. runtime, thresh = 1.0 pixel

Results – Scalability

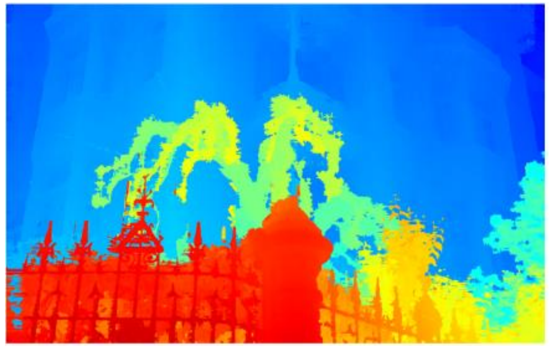
- With increasing disparity range



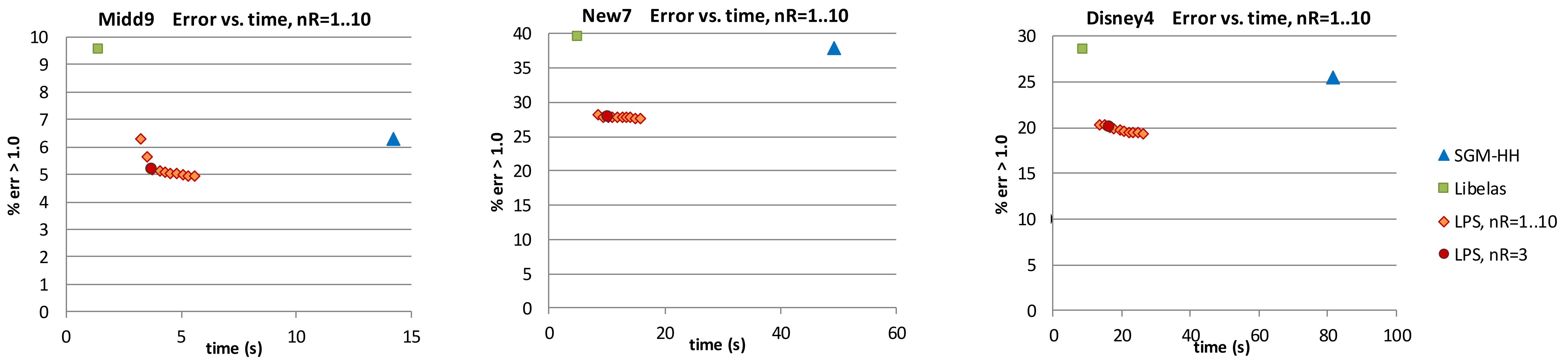
Runtime



Error



Results – more proposals



Avg. error vs. number of rounds

LPS – Summary

- Plane proposals from matched features
- Refine into surfaces using LPS
- Propagate promising planes
- Final pixel assignment to surfaces using global optimization

LPS – Benefits

- Doesn’t explore full search space
- Runtime independent of disparity range
- No fronto-parallel bias
- Excellent recovery of slanted surfaces, even with weak texture
- Easy integration of vertical disparity correction

LPS – Limitations

- Can miss surfaces if not among initial proposals
- Cannot handle completely untextured surfaces
- Need “stopping criterion” for proposal generation
- No occlusion reasoning

Future work

- Other types of proposals
  - Hierarchical (coarse to fine)
  - Line, edge features
  - New proposals via "residual analysis"
- Add color models and occlusion reasoning into final pixel assignment
- Better modeling of calibration errors

Promising directions

- Moving away from monolithic optimization
- Global pixel-to-surface assignment (like segmentation; not matching)
- Local stereo matching for proposal generation
- Residual analysis to guide additional search

Surface-based stereo matching

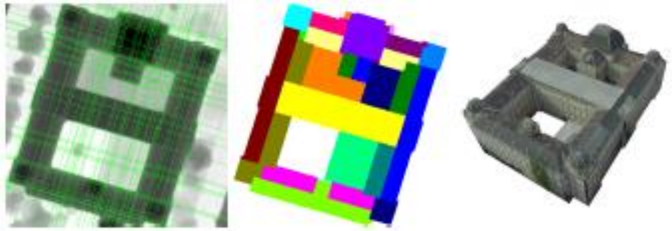
- Piecewise planar stereo

Birchfield and Tomasi 2001

Furukawa et al. 2008

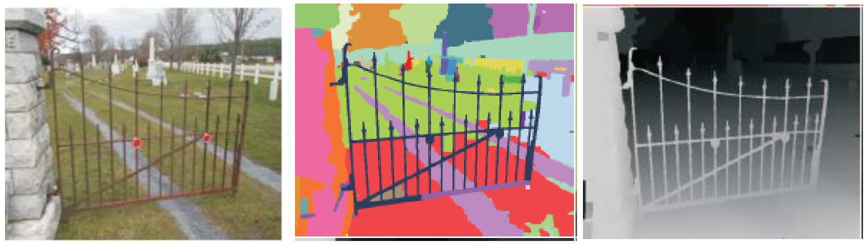
Sinha et al. 2009

- Surface stereo



Zebedin et al. 2008

Gallup et al. 2010



Bleyer et al. 2010, 2011

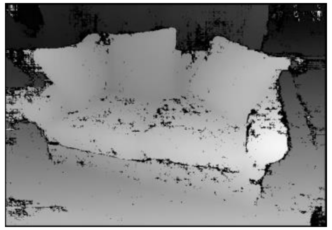
Multiple View Object Cosegmentation using Appearance and Stereo

Adarsh Kowdle, Sudipta N. Sinha and Rick Szeliski (ECCV 2012)

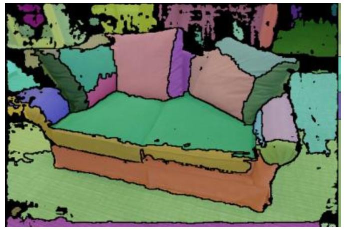
Input images

Piecewise planar stereo

- Learn per-plane local color models
- Combine stereo and color cues
- Accurate occlusion boundaries



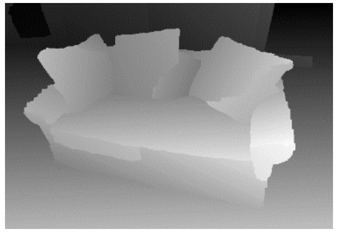
Stereo matching



Plane hypotheses

Plane labels

+

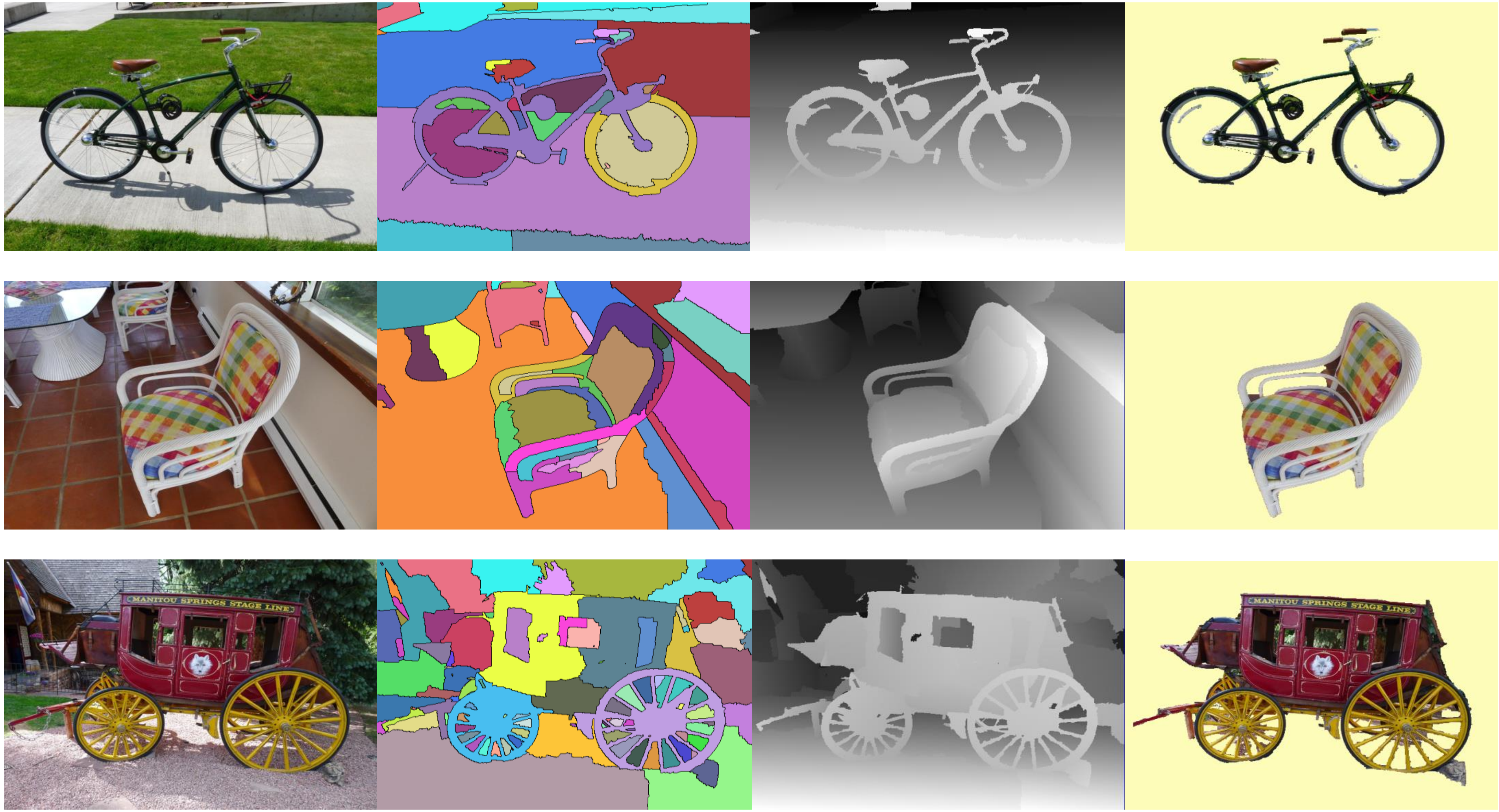


Depth map

Using multiple views, infer what constitutes the foreground *object*

Multiple View Object Cosegmentation using Appearance and Stereo

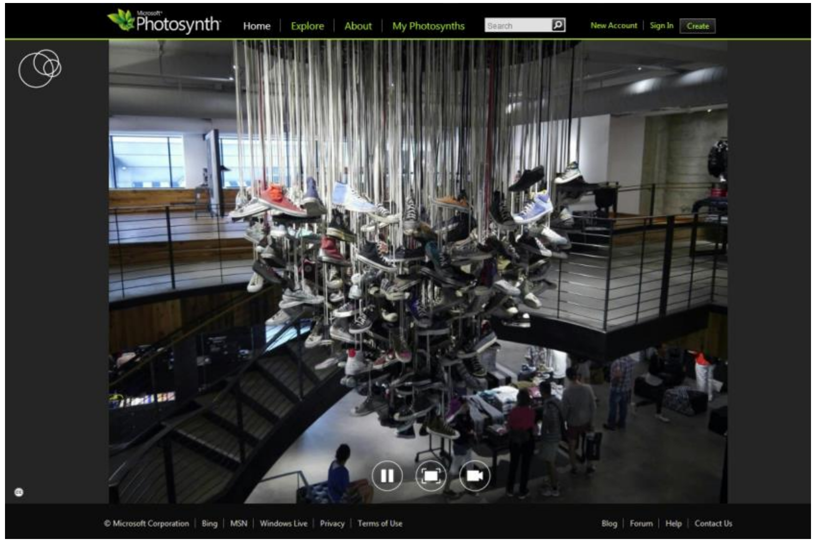
Adarsh Kowdle, Sudipta N. Sinha and Rick Szeliski (ECCV 2012)

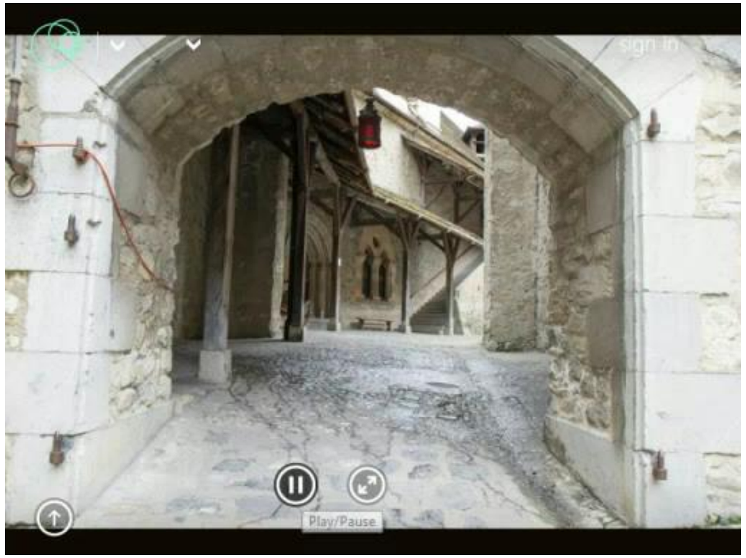


Application:
Image-based Rendering

Photosynth 2 (www.photosynth.net)

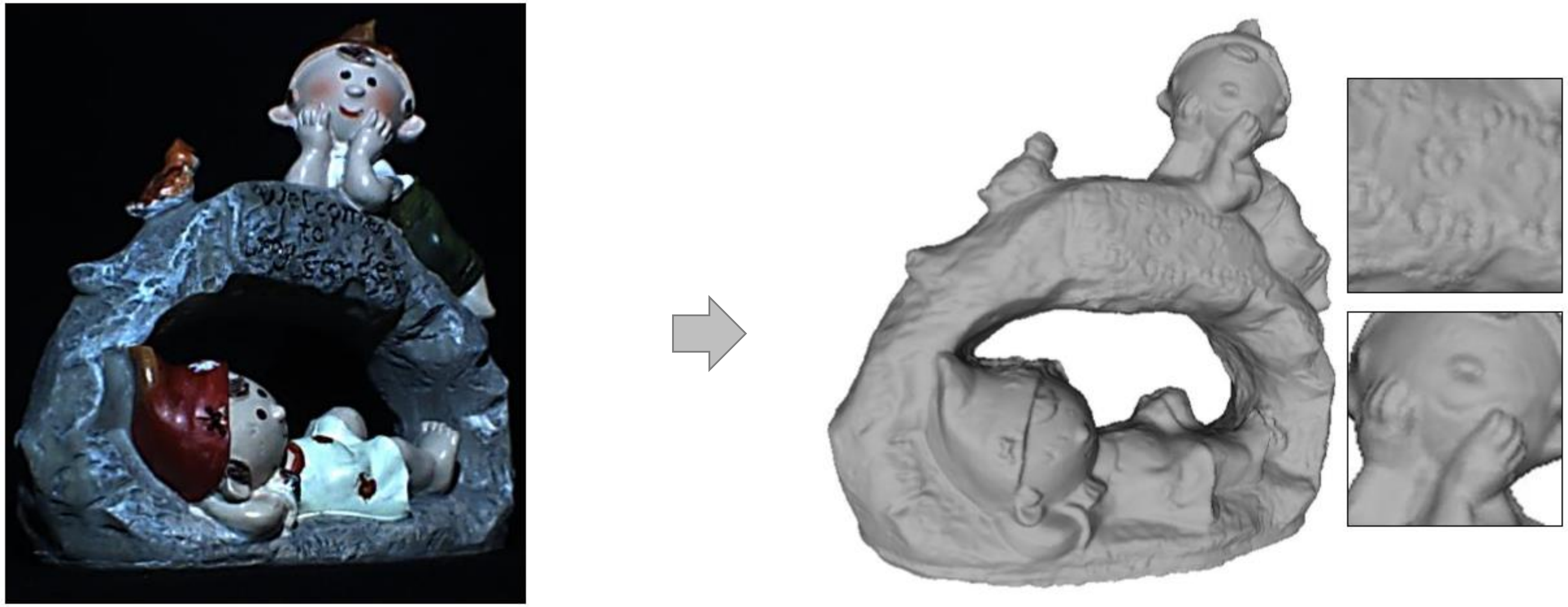
- Capture the world in 3D;
- Novel view synthesis
- Interactive viewer





Multiview Photometric Stereo using Planar Mesh Parameterization

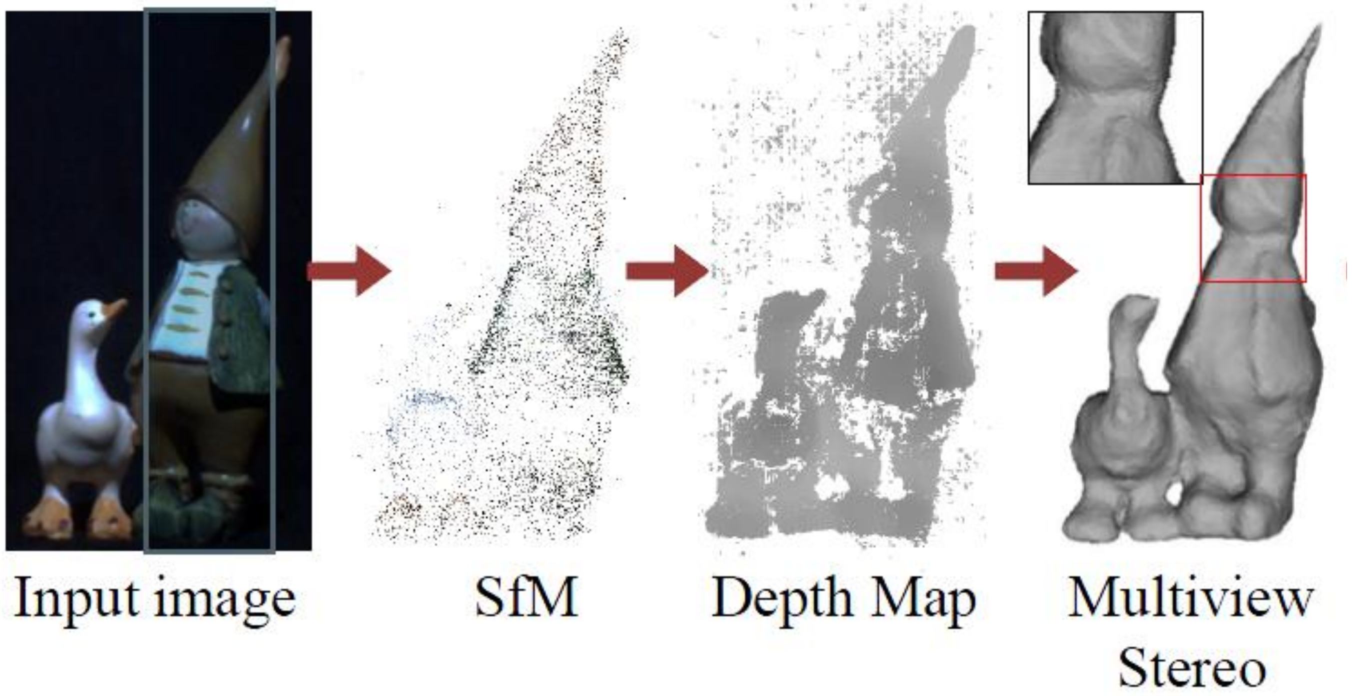
Jaesik Park, Sudipta N. Sinha, Yasuyuki Matsushita, Yu-Wing Tai and In So Kweon (ICCV 2013)



- Automatic 3D reconstruction from RGB images
- Multi-view stereo gives coarse shape
- Multi-view Photometric stereo refines shape

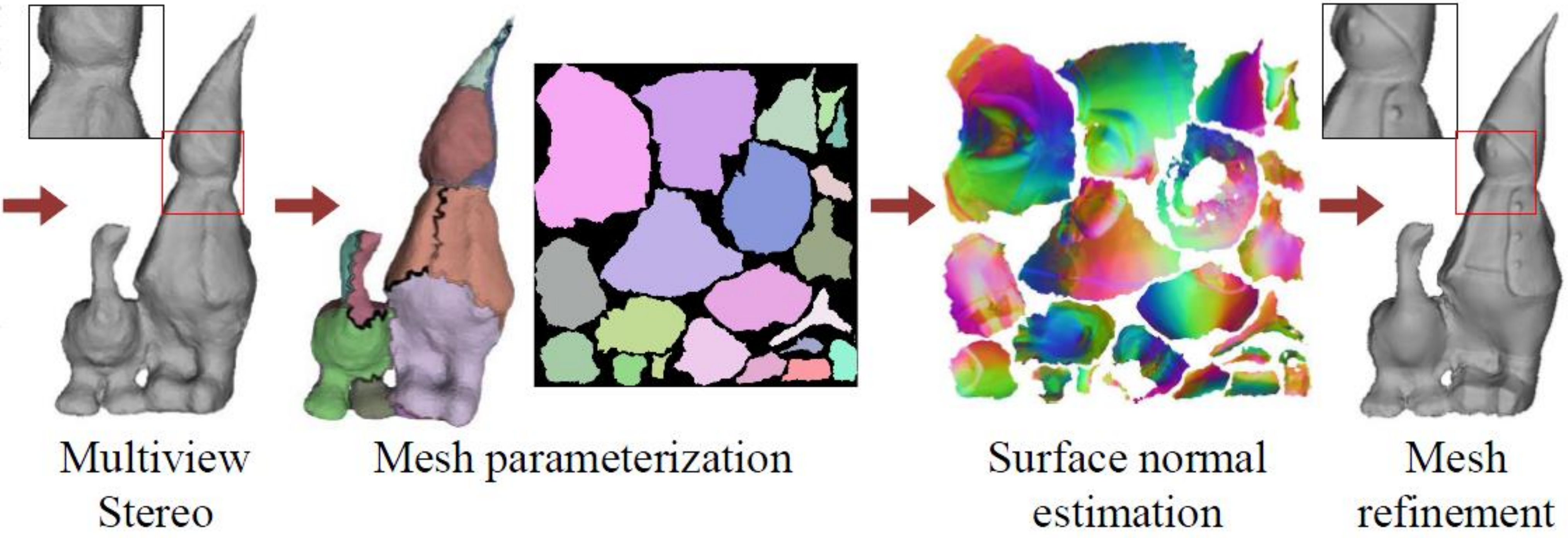
Multiview Photometric Stereo using Planar Mesh Parameterization

Jaesik Park, Sudipta N. Sinha, Yasuyuki Matsushita, Yu-Wing Tai and In So Kweon (ICCV 2013)



Multiview Photometric Stereo using Planar Mesh Parameterization

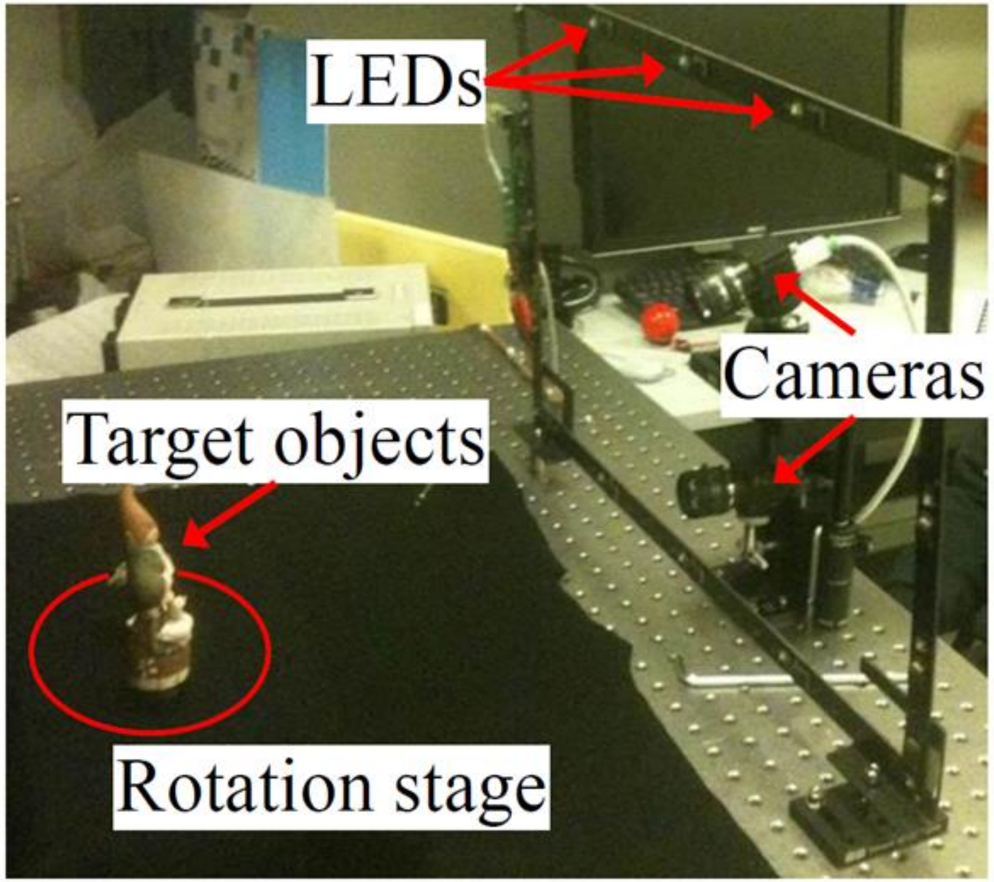
Jaesik Park, Sudipta N. Sinha, Yasuyuki Matsushita, Yu-Wing Tai and In So Kweon (ICCV 2013)



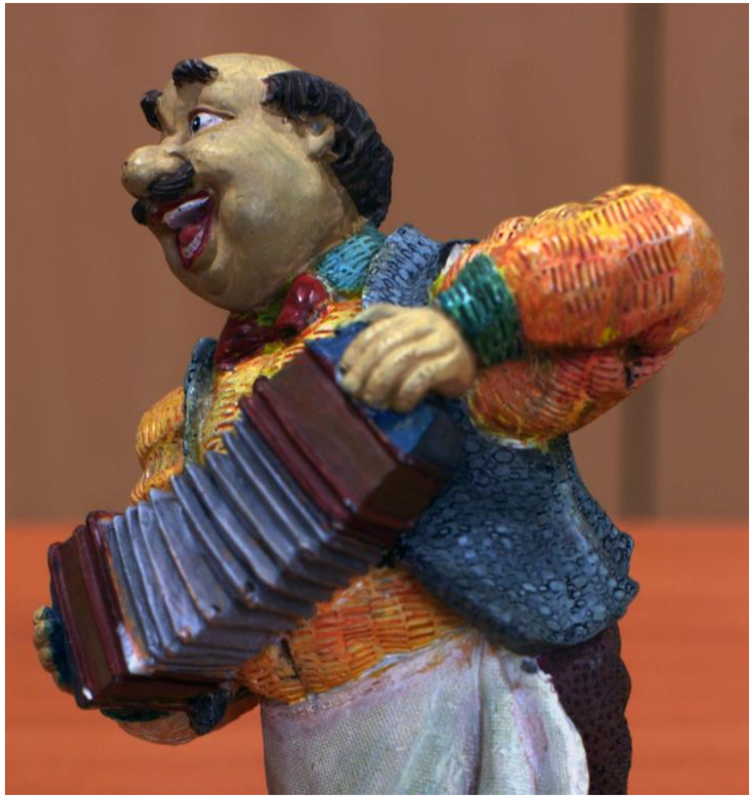
Multiview Photometric Stereo using Planar Mesh Parameterization

Jaesik Park, Sudipta N. Sinha, Yasuyuki Matsushita, Yu-Wing Tai and In So Kweon (ICCV 2013)

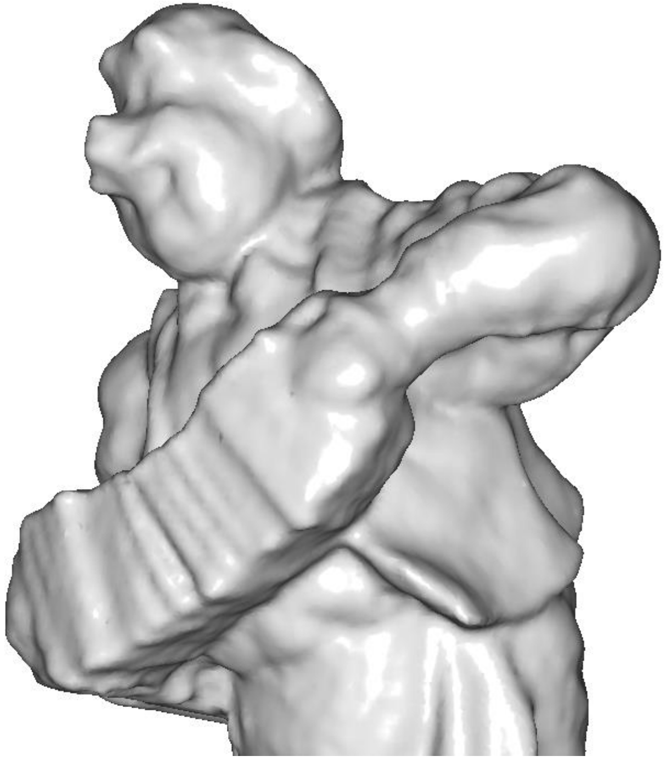
Acquisition Setup



Multiview Photometric Stereo using Planar Mesh Parameterization



Input image

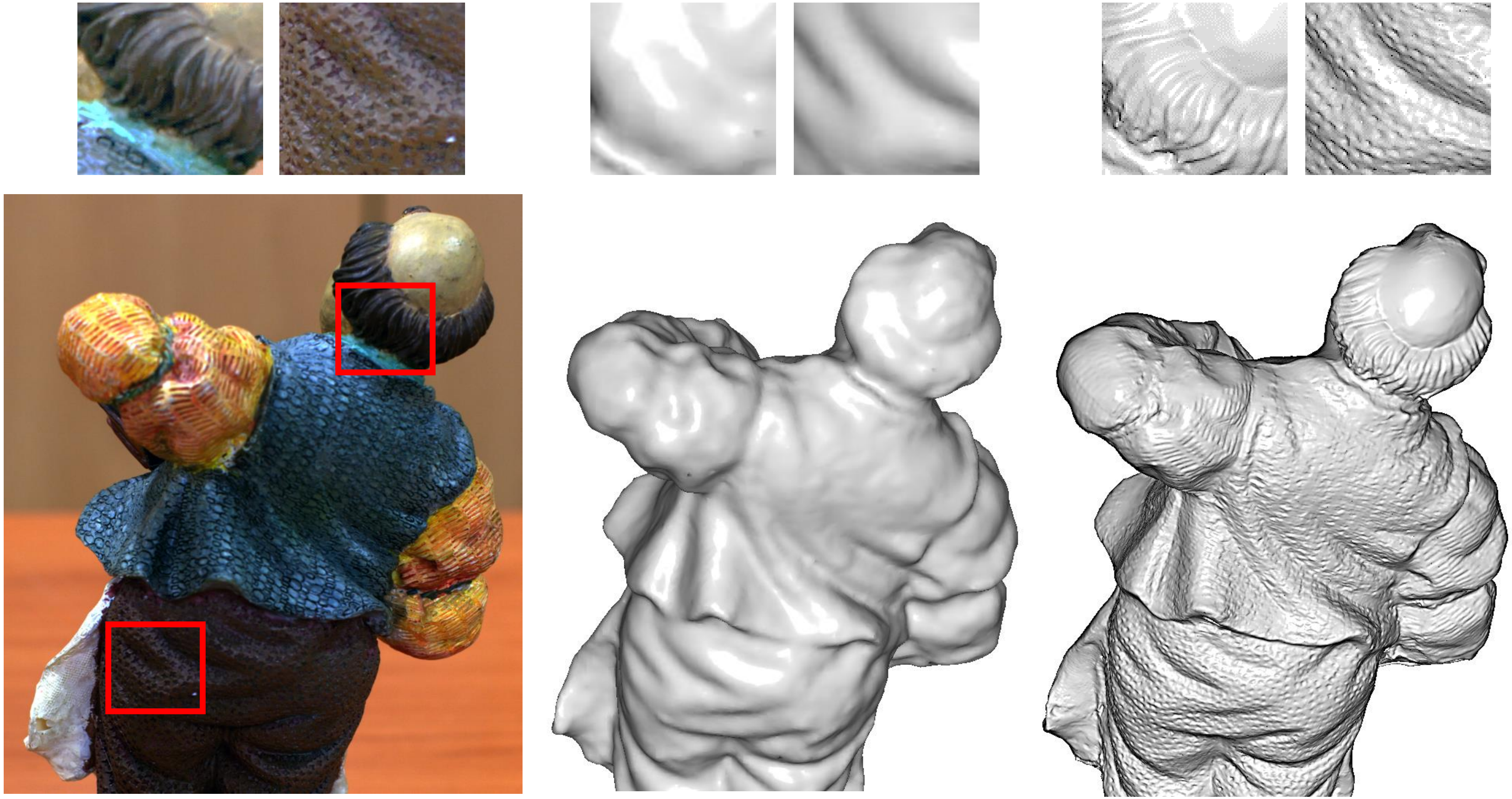


Base Mesh (Multiview Stereo)

Final 3D model

Fine geometric details recovered from hi-res images (10+ MP)

Multiview Photometric Stereo using Planar Mesh Parameterization



Input image | Base Mesh (Multiview Stereo) | Final 3D model

Conclusion

- New directions for hi-res stereo matching
- New datasets and benchmarks are coming !
- Local Plane Sweep stereo

  - avoid exploring the whole DSI

  - moving away from monolithic optimization
- Surface-based stereo matching

  - robust; allows a range of priors to be incorporated
- Combining multi-view stereo and photometric stereo

Acknowledgements

Daniel Scharstein, Rick Szeliski, Adarsh Kowdle,

Michael Bleyer, Carsten Rother, Pushmeet Kohli,

Jaesik Park, Yasuyuki Matsushita, Yu-Wing Tai, In So Kweon

Thanks!