

Deep Stereo

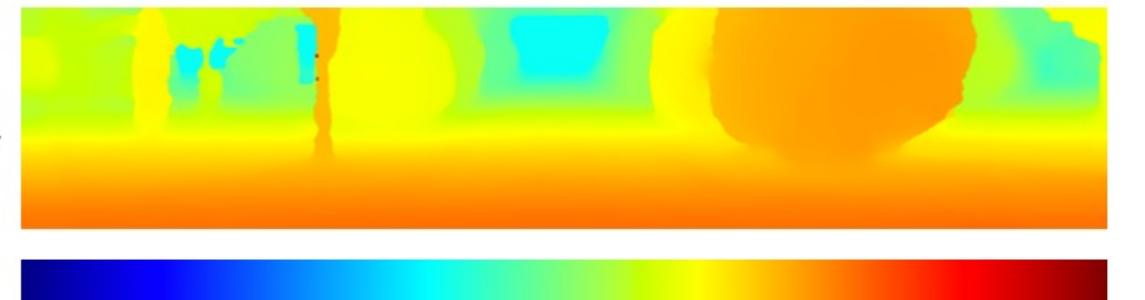
Frank Dellaert, x476 Fall 2021

Left input image



Right input image

Output disparity map



90 m

20 m

1.7 m

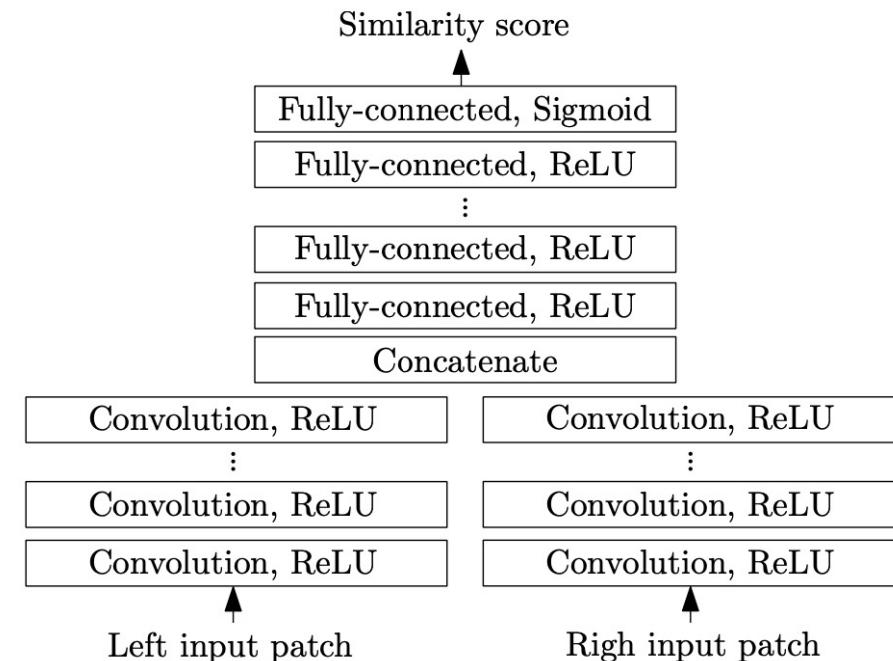
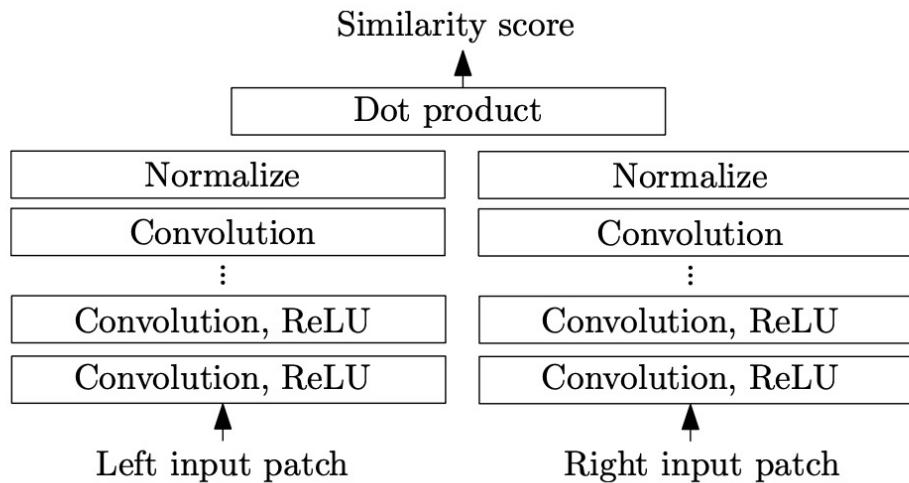
Image from Žbontar & LeCun, 2016

First Idea: matching costs

Stereo Matching by Training a Convolutional Neural Network to Compare Image Patches

Jure Žbontar*

Yann LeCun[†]



- Two versions: fast, and accurate
 - Tries to make distance between positive examples (matched patches) and negative (incorrectly matched patches) large

Results

- On KITTI (2012) dataset, in October 2015
- percentage of misclassified pixels

Rank	Method		Setting	Error	Runtime
1	MC-CNN-acrt	Accurate architecture		2.43	67
2	Displets	Güney and Geiger (2015)		2.47	265
3	MC-CNN	Žbontar and LeCun (2015)		2.61	100
4	PRSM	Vogel et al. (2015)	F, MV	2.78	300
	MC-CNN-fst	Fast architecture		2.82	0.8
5	SPS-StFl	Yamaguchi et al. (2014)	F, MS	2.83	35
6	VC-SF	Vogel et al. (2014)	F, MV	3.05	300
7	Deep Embed	Chen et al. (2015)		3.10	3
8	JSOSM	Unpublished work		3.15	105
9	OSF	Menze and Geiger (2015)	F	3.28	3000
10	CoR	Chakrabarti et al. (2015)		3.30	6

Similar idea:

Efficient Deep Learning for Stereo Matching

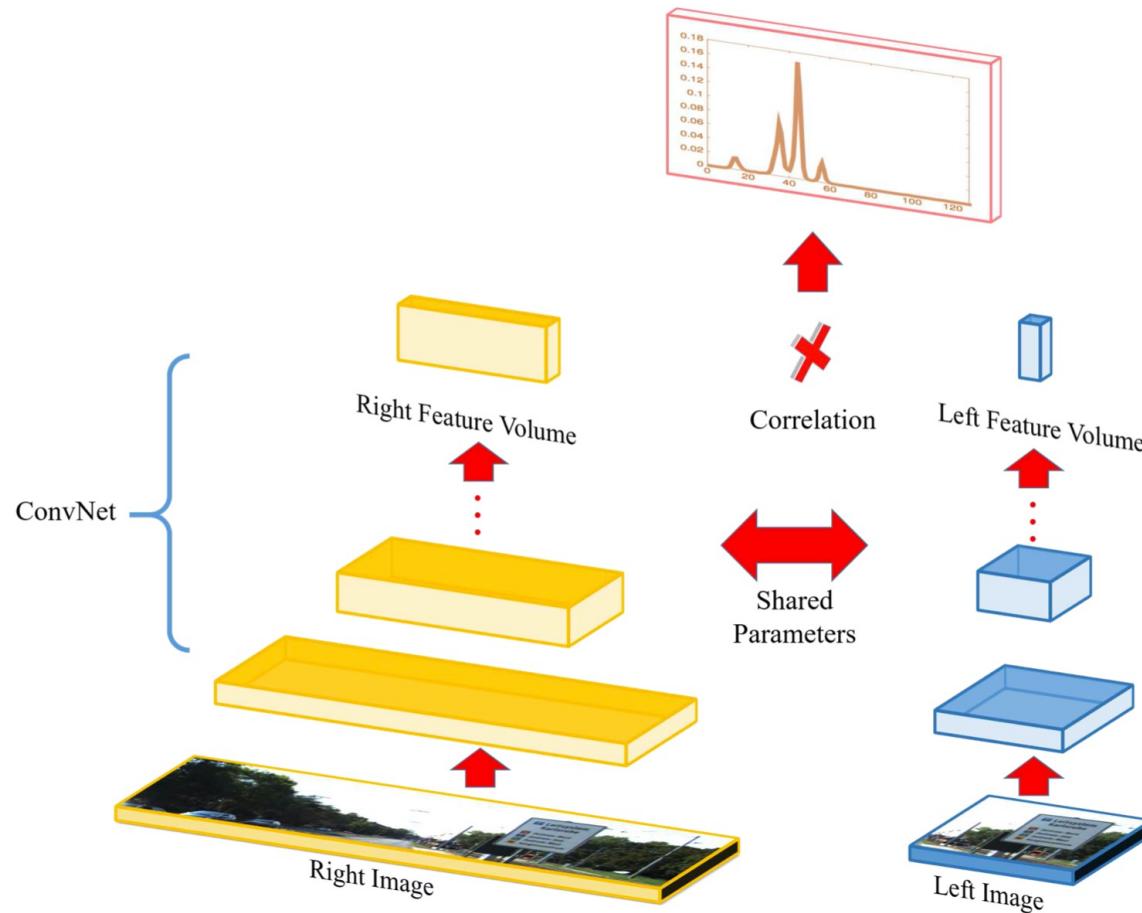
CVPR 2016

Wenjie Luo

Alexander G. Schwing

Raquel Urtasun

Department of Computer Science, University of Toronto

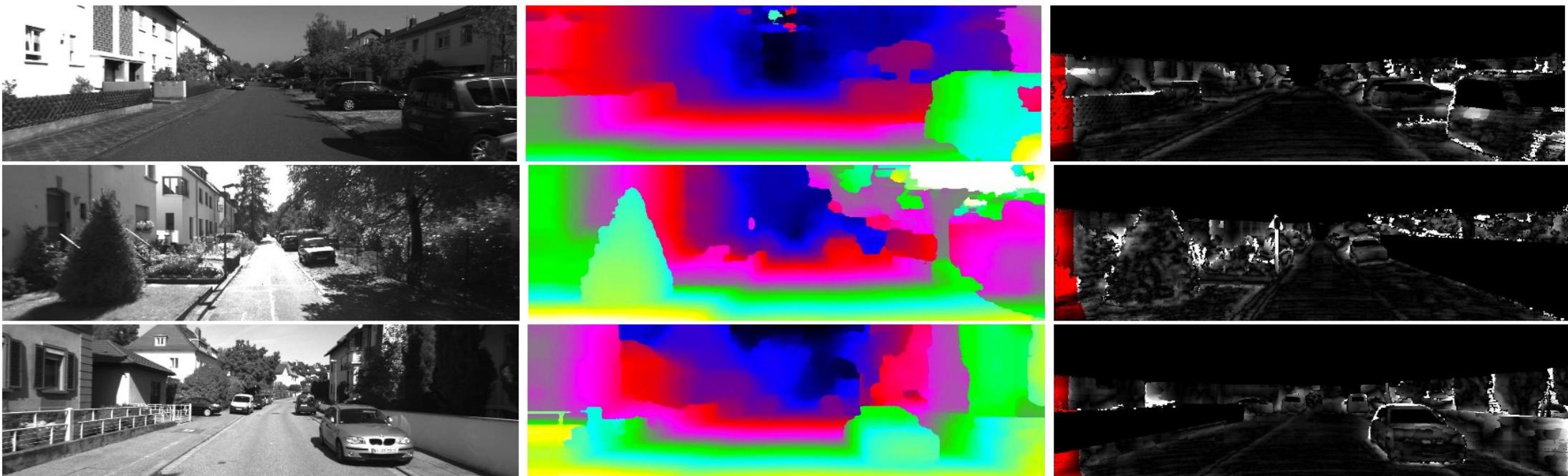


- CNN for feature representation, shared parameters
- Probability density over disparities

Results (percentage over threshold)

	> 2 pixels		> 3 pixels		> 4 pixels		> 5 pixels		End-Point		Runtime (s)
	Non-Occ	All	Non-Occ	All	Non-Occ	All	Non-Occ	All	Non-Occ	All	
StereoSLIC [26]	5.76	7.20	3.92	5.11	3.04	4.04	2.49	3.33	0.9 px	1.0 px	2.3
PCBP-SS [26]	5.19	6.75	3.40	4.72	2.62	3.75	2.18	3.15	0.8 px	1.0 px	300
SPS-st [27]	4.98	6.28	3.39	4.41	2.72	3.52	2.33	3.00	0.9 px	1.0 px	2
Deep Embed [7]	5.05	6.47	3.10	4.24	2.32	3.25	1.92	2.68	0.9 px	1.1 px	3
MC-CNN-acrt [30]	3.90	5.45	2.43	3.63	1.90	2.85	1.64	2.39	0.7 px	0.9 px	67
Displets v2 [12]	3.43	4.46	2.37	3.09	1.97	2.52	1.72	2.17	0.7 px	0.8 px	265
Ours(19)	4.98	6.51	3.07	4.29	2.39	3.36	2.03	2.82	0.8 px	1.0 px	0.7

Table 3: Comparison to stereo state-of-the-art on the test set of the KITTI 2012 benchmark.



Cost Volume Aggregation

End-to-End Learning of Geometry and Context for Deep Stereo Regression

ICCV 2017

Alex Kendall

Hayk Martirosyan

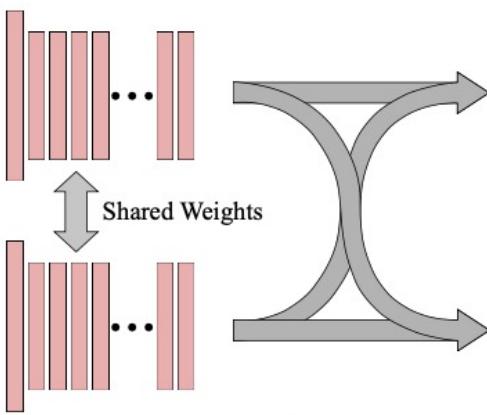
Saumitro Dasgupta

Peter Henry

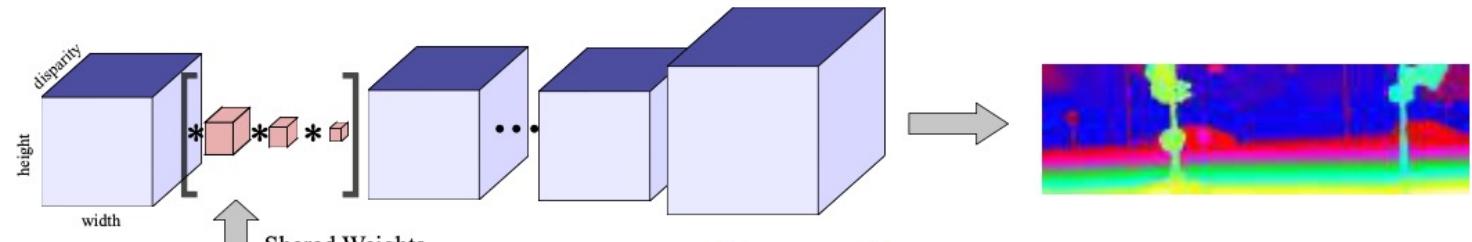
Ryan Kennedy

Abraham Bachrach

Adam Bry



Skydio Research



Input Stereo Images

2D Convolution

Cost Volume

Multi-Scale 3D Convolution

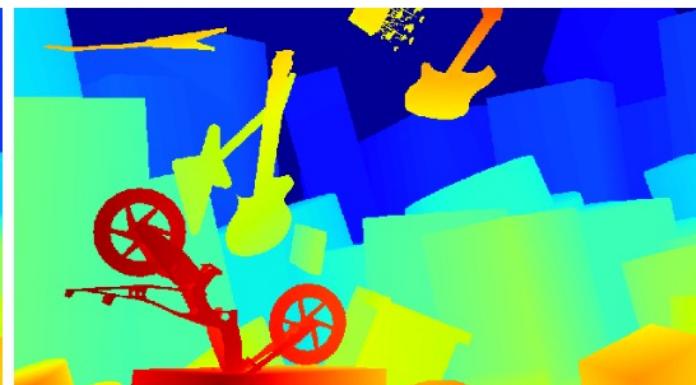
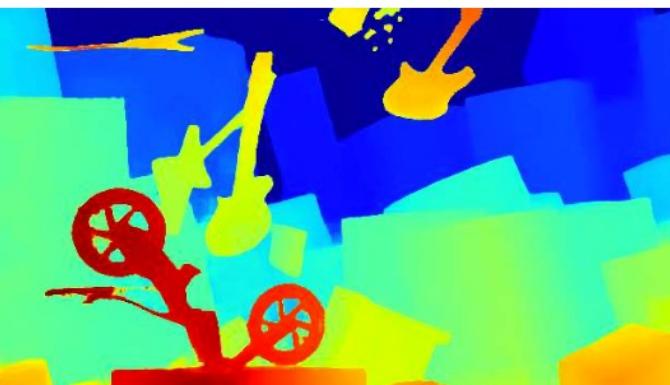
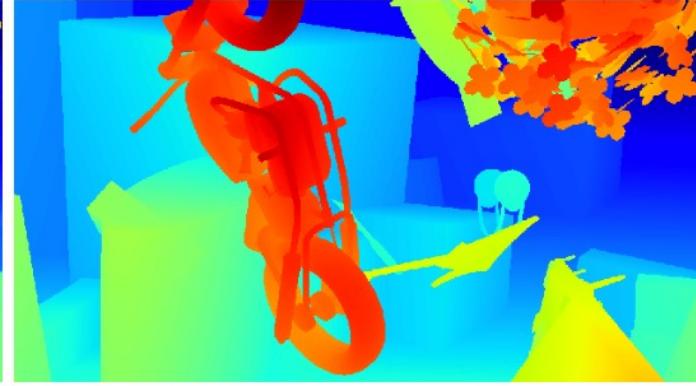
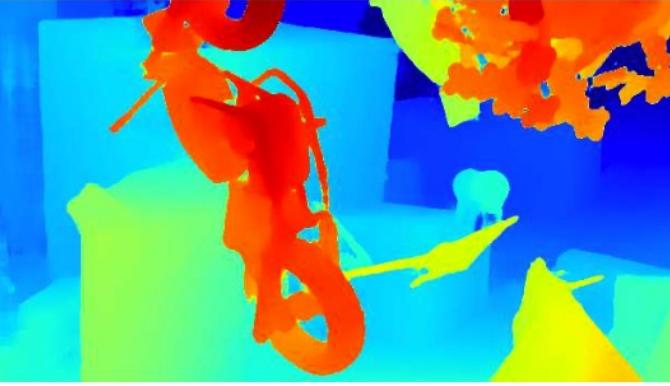
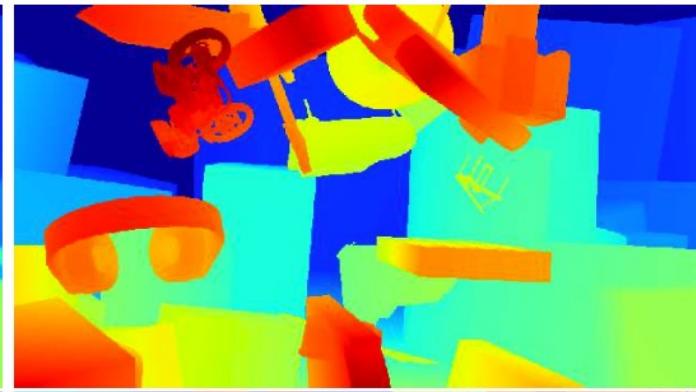
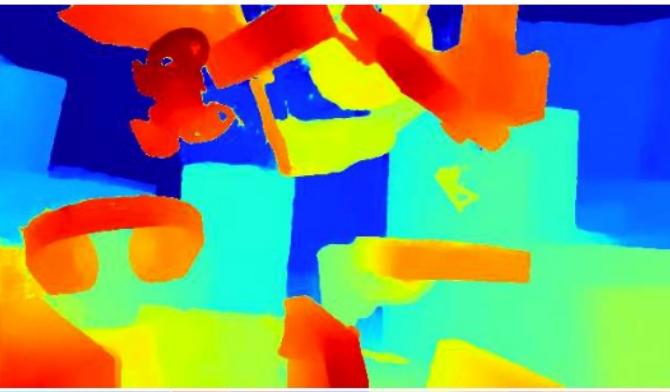
3D Deconvolution

Soft ArgMax

Disparities

- CNN for features, shared weights, then 3D convolutions

Results on Scene Flow Dataset (from CVPR 2016)



Pyramid Stereo Matching Network

CVPR 2018

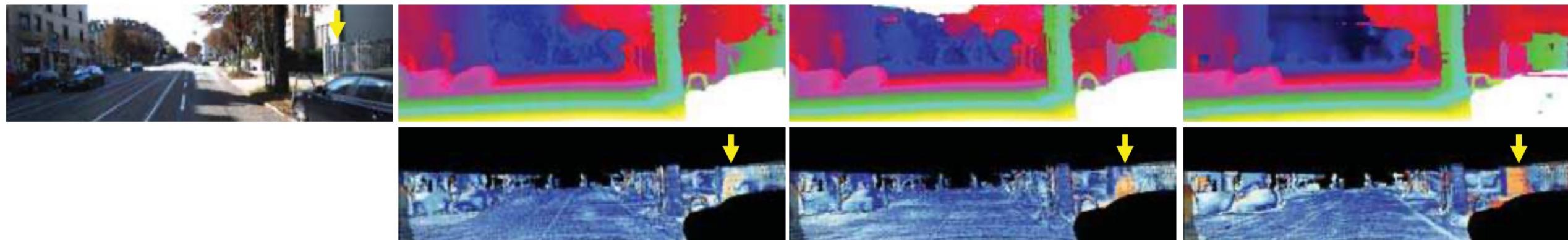
Jia-Ren Chang

Yong-Sheng Chen

Department of Computer Science, National Chiao Tung University, Taiwan

- Architectural improvements: spatial pyramid pooling
- Results on KITTI 2015, March 2018 leaderboard:

Rank	Method	All (%)			Noc (%)			Runtime (s)
		D1-bg	D1-fg	D1-all	D1-bg	D1-fg	D1-all	
1	PSMNet (ours)	1.86	4.62	2.32	1.71	4.31	2.14	0.41
3	iResNet-i2e2 [14]	2.14	3.45	2.36	1.94	3.20	2.15	0.22
6	iResNet [14]	2.35	3.23	2.50	2.15	2.55	2.22	0.12
8	CRL [21]	2.48	3.59	2.67	2.32	3.12	2.45	0.47
11	GC-Net [13]	2.21	6.16	2.87	2.02	5.58	2.61	0.90



Hierarchical Deep Stereo Matching on High-resolution Images

CVPR 2019

Gengshan Yang^{1,*}, Joshua Manela², Michael Happold², Deva Ramanan^{1,2}
¹Carnegie Mellon University, ²Argo AI

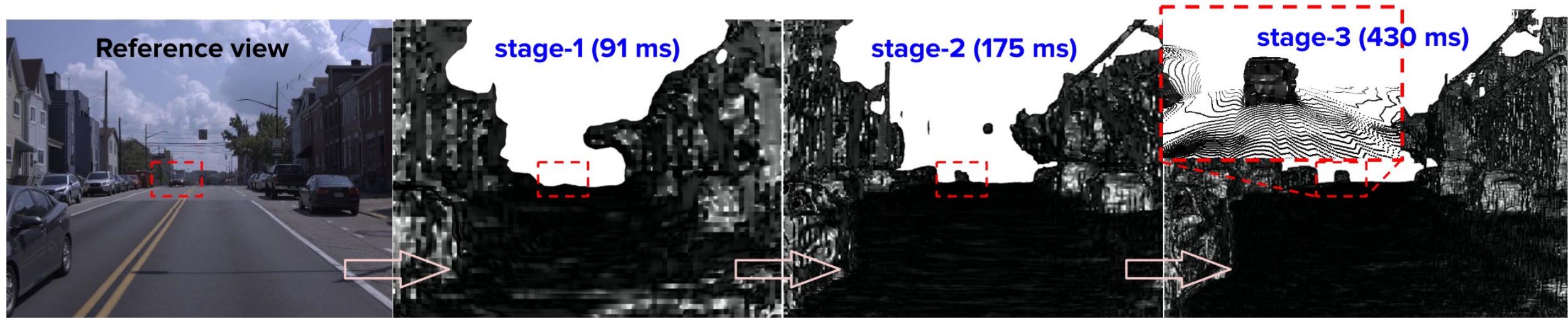
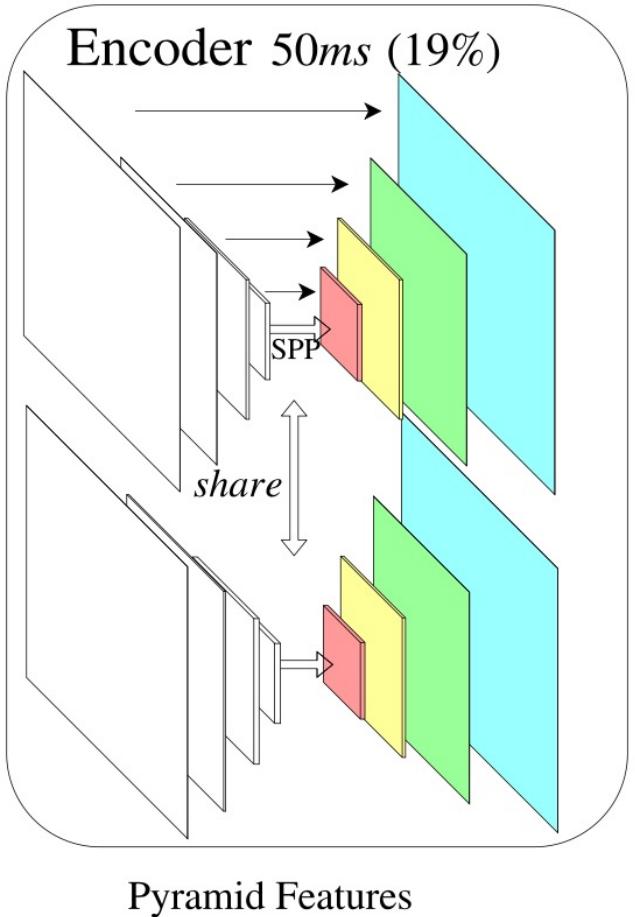


Figure 1: Illustration of on-demand depth sensing with a coarse-to-fine hierarchy on the proposed dataset. Our method (HSM) captures the coarse layout of the scene in 91 milliseconds, finds the far-away car (shown in the red box) in 175 ms, and recovers the details of the car given extra 255 ms.

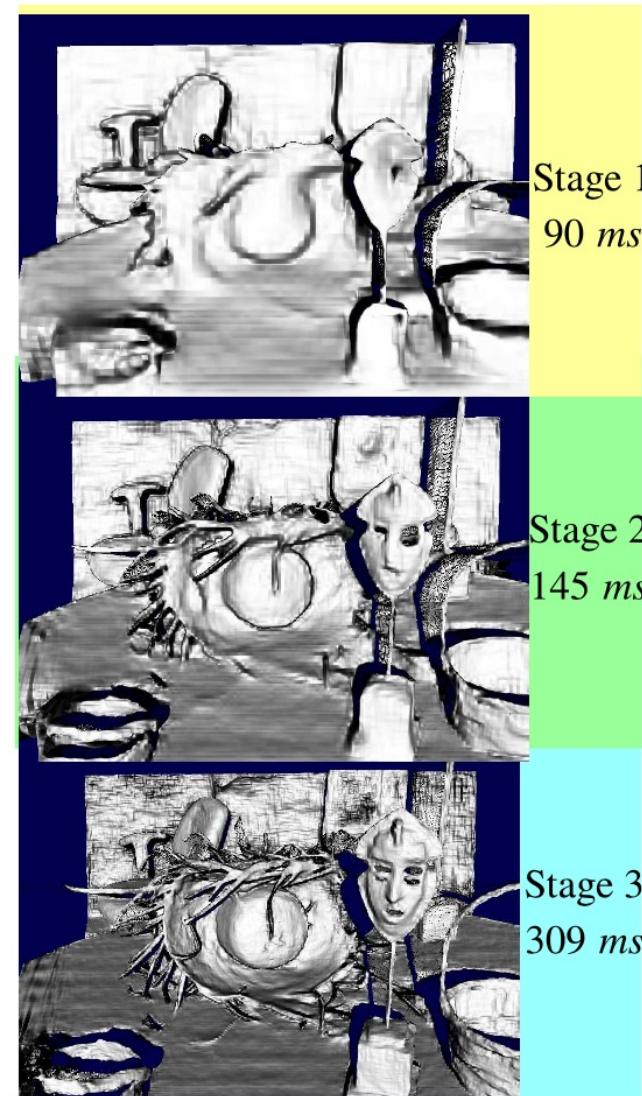
- High-resolution for self-driving: $Z=f B/Z$, increase f !

Yang18cvpr architecture figure:



$$\frac{H \times W}{\{8, 16, 32, 64\}} \times C_k$$

$$\frac{H \times W \times D_k}{\{8, 16, 32, 64\}}$$

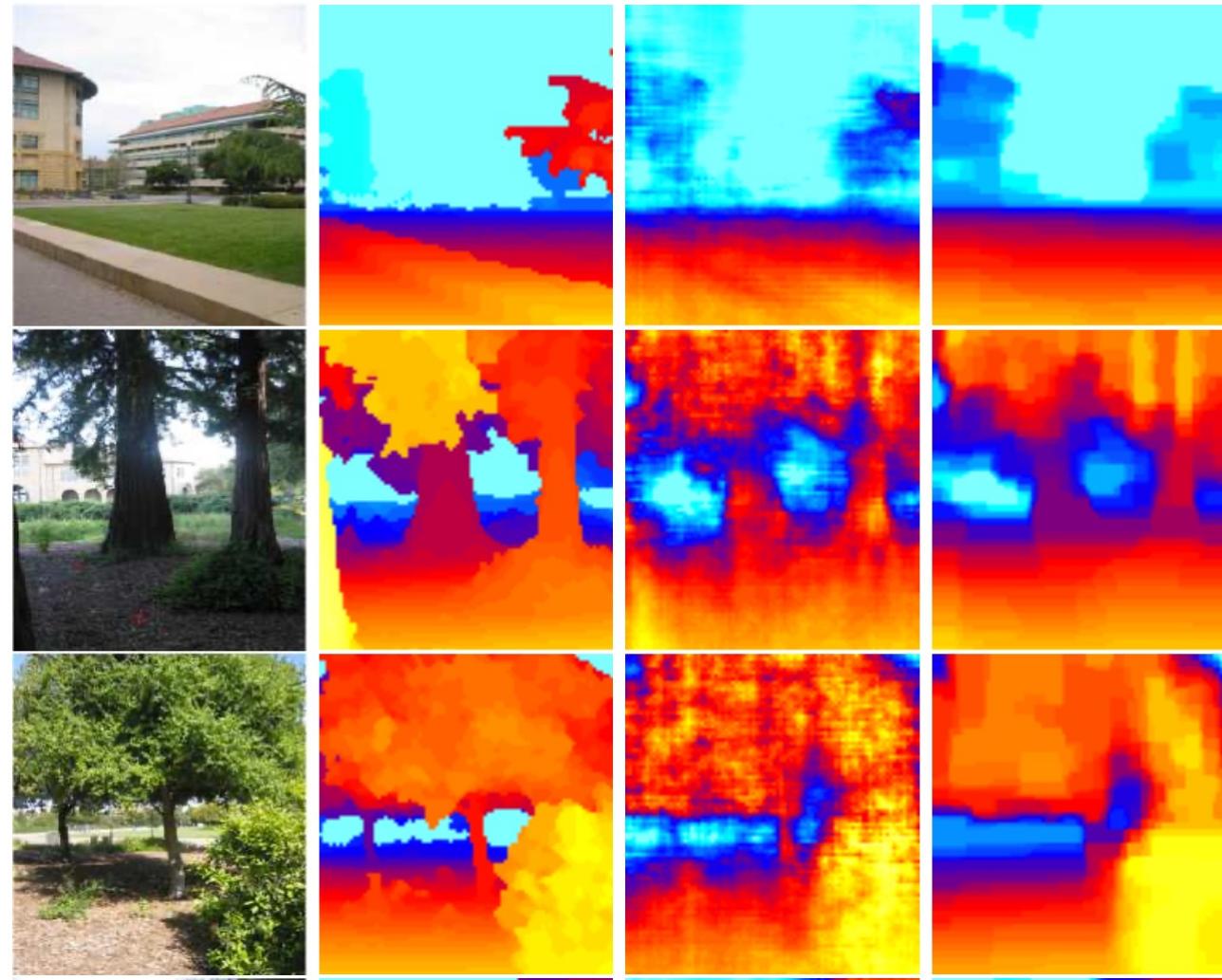


Learning Depth from Single Monocular Images

NIPS 2005 (!)

Ashutosh Saxena, Sung H. Chung, and Andrew Y. Ng

- A whole different beast: monocular depth
- Not deep: Markov random field (MRF)
- Learns a relatively small number of parameters



Unsupervised Monocular Depth Estimation with Left-Right Consistency

CVPR 2017

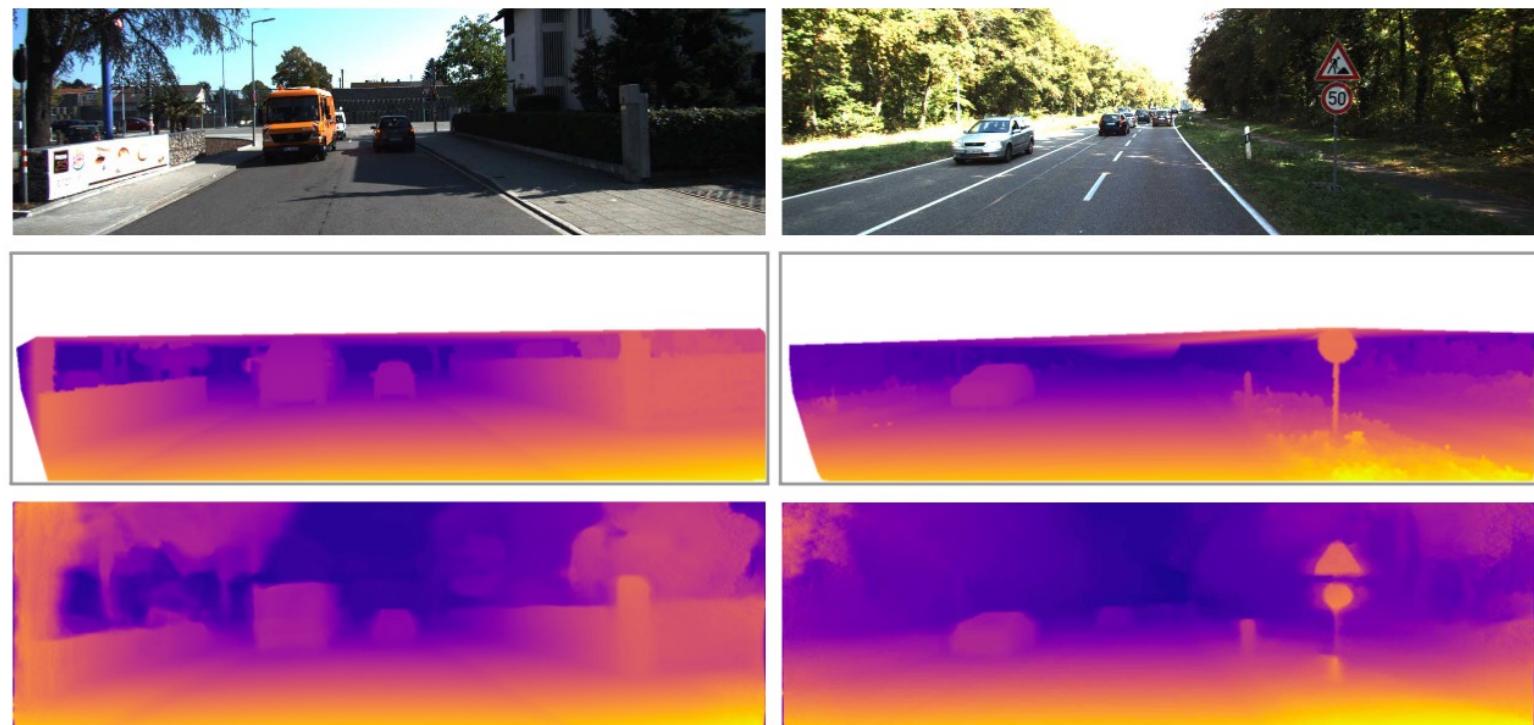
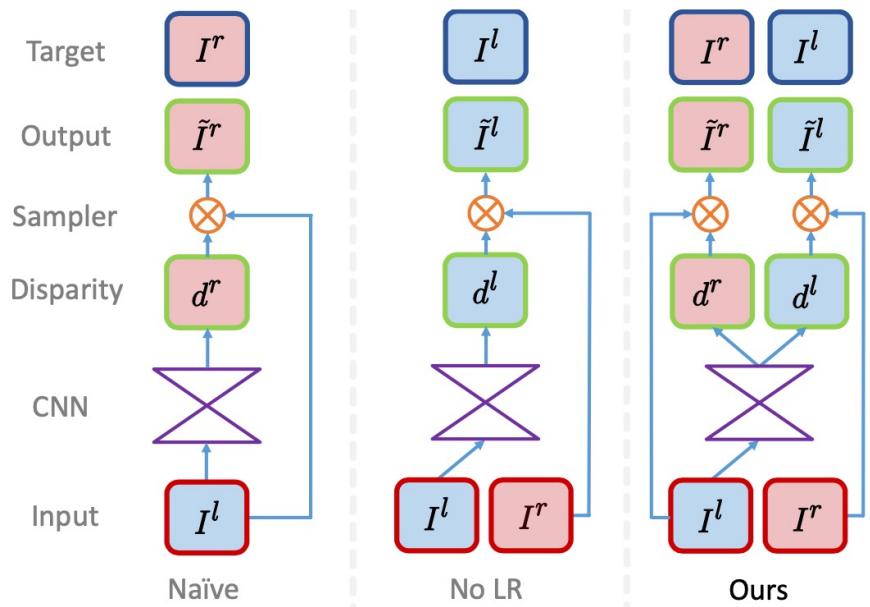
Clément Godard

Oisin Mac Aodha

Gabriel J. Brostow

University College London

<http://visual.cs.ucl.ac.uk/pubs/monoDepth/>



- LR- consistency
- Unsupervised monocular depth