Depth From Stereo Camera

Enabling Machines to see as Human

Compiled by

Pankaj Kumar Bajpai

Samsung R&D Institute India, Bangalore

Dec 14, 2018

SAMSUNG

© Samsung 2018

Contents

© Samsung 2018

An Introduction to

STEREO VISION

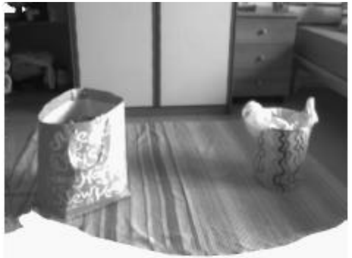
SAMSUNG

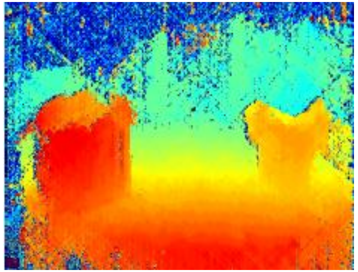
© Samsung 2018

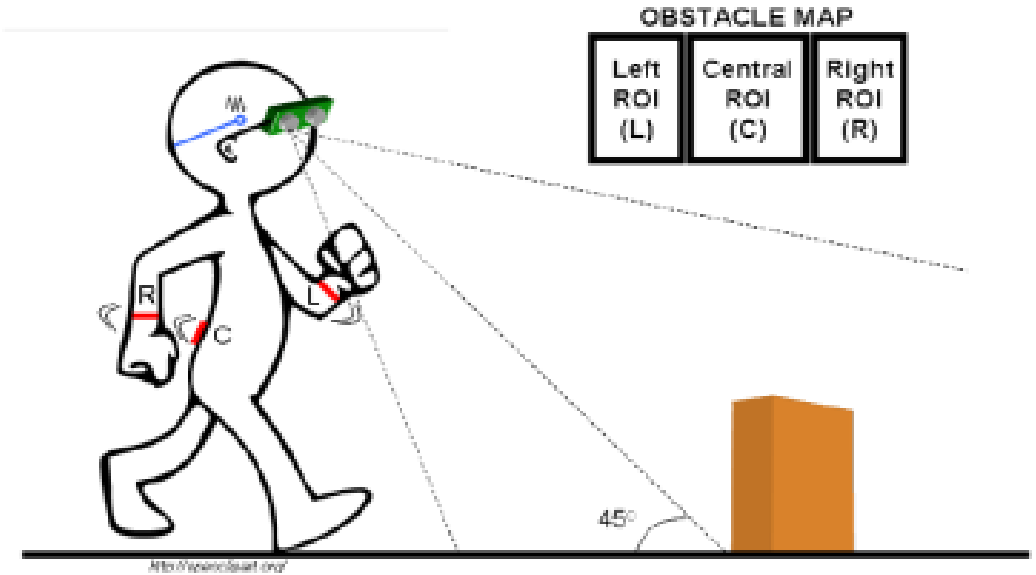
Why ? Importance of Depth

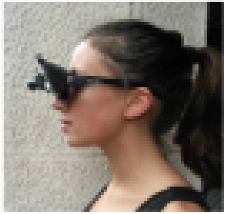
© Samsung 2018

Autonomous Robot/
vehicle navigation









What ? Stereo Vision

SAMSUNG
© Samsung 2018

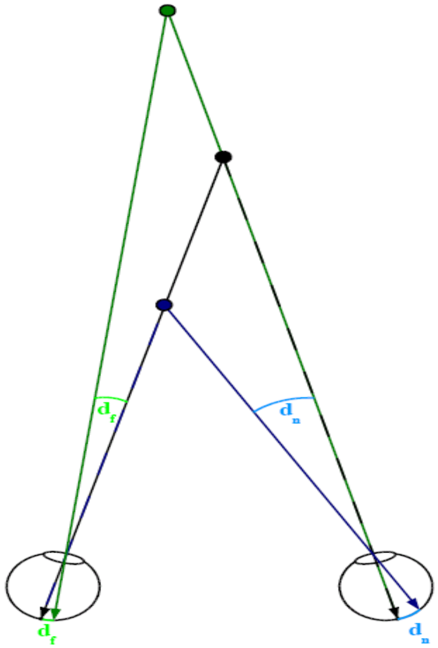
➢ Close one of your eyes and try complicated tasks like tossing an object and catching it. Ask yourselves the following questions

- Can I perceive depth with one eye closed?
- If so, what cues does my eye use?
- Will it work well under all circumstances (like playing sports)?

Stereo Vision ⇔ **Human Eye**

What ? Stereo Vision aka Binocular Vision

© Samsung 2018



Two images of a stereoscopic photograph. The difference between the two images, such as the distances between the front cactus and the window in the two views, creates retinal disparity. This creates a perception of depth when (a) the left image is viewed by the left eye and (b) the right image is viewed by the right eye.

- Binocular **Disparity** : Relative 2D displacement of the image of the same point in space when projected on two different focal planes (i.e. two different eyes)
  - Objects closer to eye ➔ higher retinal disparity
  - Disparity inversely proportional to depth

Depth Perception : Other Modalities

SAMSUNG

© Samsung 2018

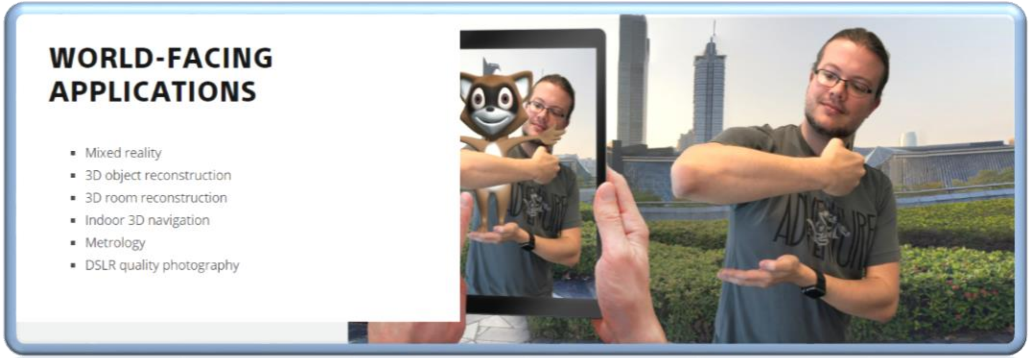
- Focus
- Atmosphere
- Perspective
- Occlusion
- Motion based
- Past learning?

© Samsung 2018

Scene analysis and 3D reconstruction

https://www.sony-depthsensing.com/Depthsense/Markets/HMD

https://www.sony-depthsensing.com/DepthSense/Markets/Automotive

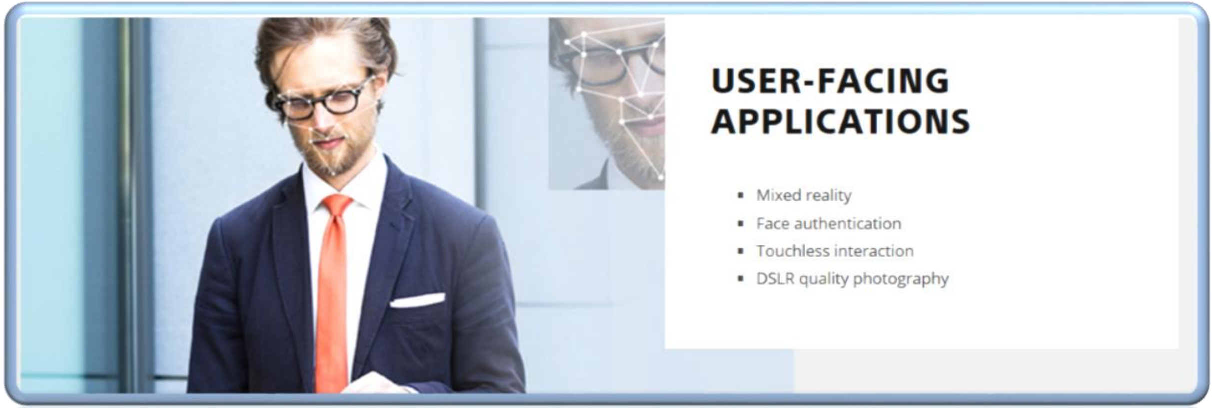


https://www.sony-depthsensing.com/Depthsense/Markets/HMD

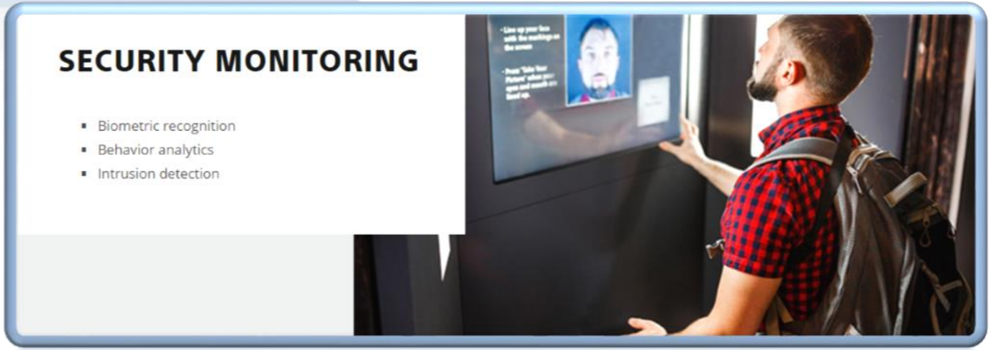
Where ? – Applications

© Samsung 2018

Face modelling



https://www.sony-depthsensing.com/Depthsense/Markets/Mobile

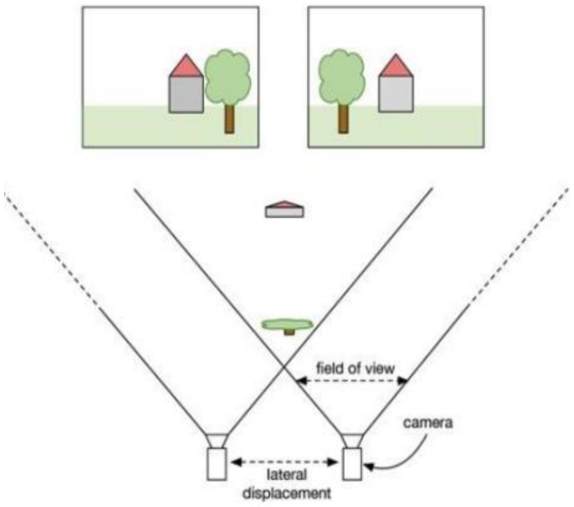


PROBLEM FORMULATION

© Samsung 2018

How ? Stereo Disparity & Depth

SAMSUNG
© Samsung 2018

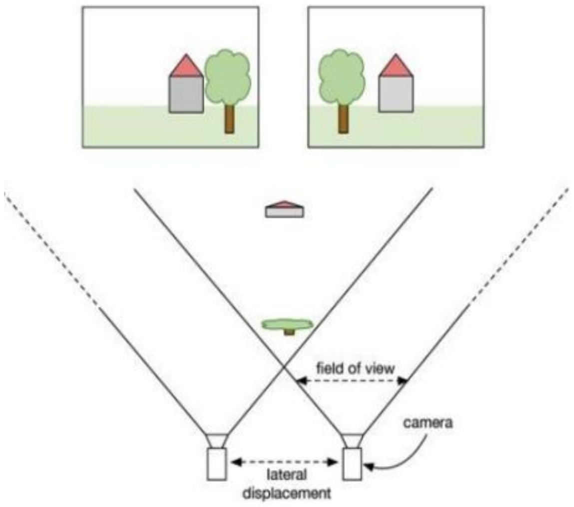


- Rectified pairs: 1D search space

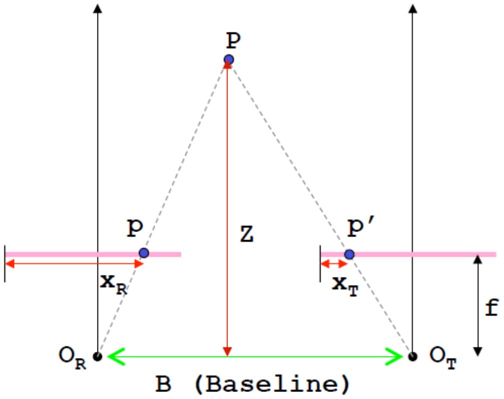
[Scharstein, D. & Szeliski, R. A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms]

How ? Stereo Disparity & Depth

SAMSUNG
© Samsung 2018



- Rectified pairs: 1D search space



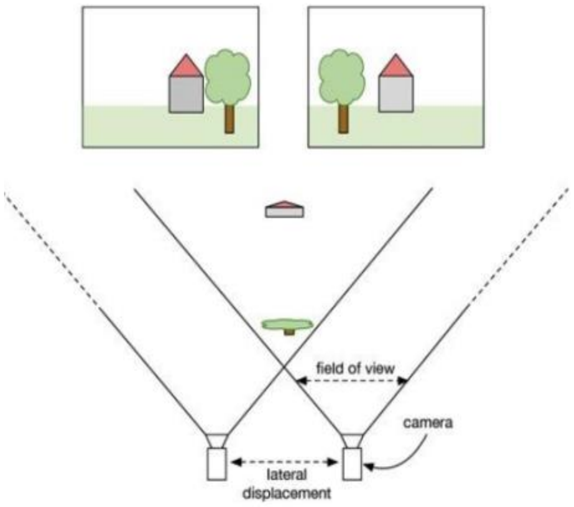
$$D(x,y) = x_R - x_T$$

$$Z(x,y) = \frac{B * f}{D(x,y)}$$

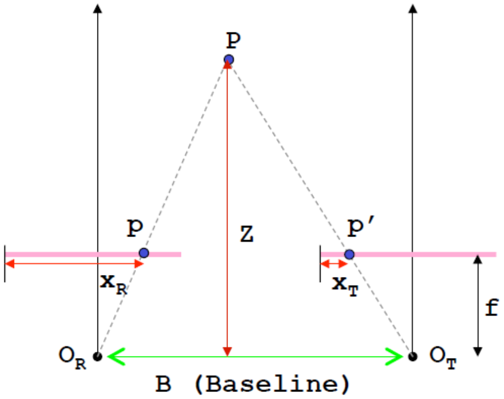
[Scharstein, D. & Szeliski, R. A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms]

How ? Stereo Disparity & Depth

SAMSUNG
© Samsung 2018



- Rectified pairs: 1D search space



$$D(x,y) = x_R - x_T$$

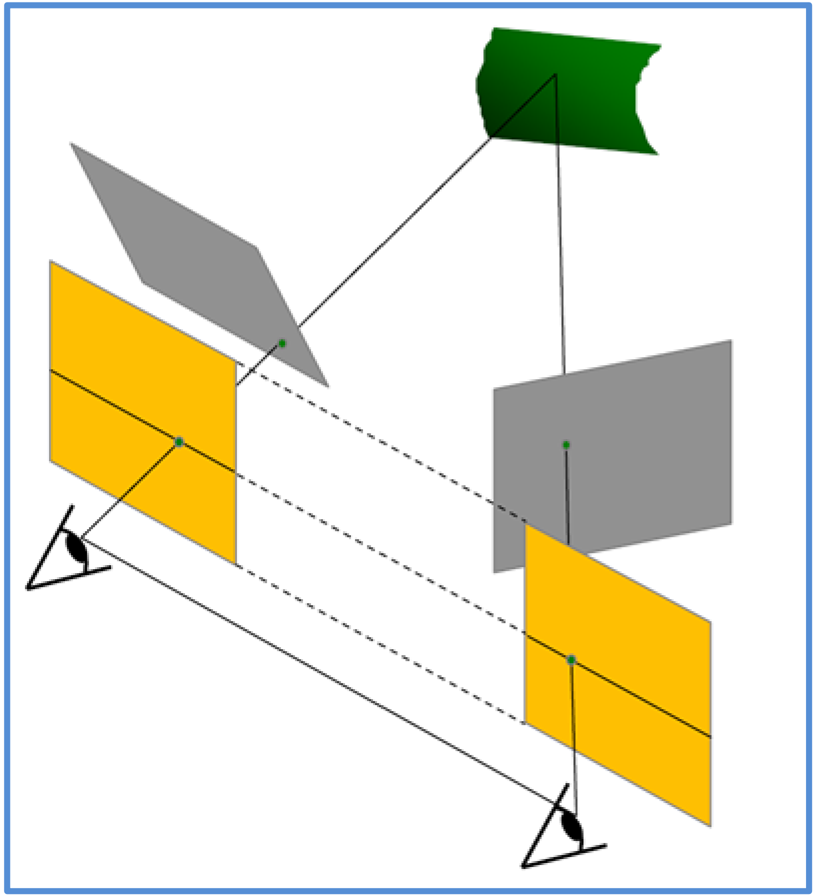
$$Z(x,y) = \frac{B * f}{D(x,y)}$$

- Assumptions
  - f, B are known (through camera calibration)
  - The epipolar lines run horizontally
  - The points p and p' are visible in both views
- Challenges
  - For every point p in left image, how to find p'?
    - ➔ Stereo correspondence problem

[Scharstein, D. & Szeliski, R. A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms]

Assumption : Rectified Stereo Image Pair

SAMSUNG
© Samsung 2018



Input to Stereo Disparity Algorithm is considered to be Rectified Stereo Image Pair

- C. Loop and Z. Zhang. Computing Rectifying Homographies for Stereo Vision. IEEE Conf. Computer Vision and Pattern Recognition, 1999.

COMPUTER VISION APPROACHES

© Samsung 2018

Computer Vision (CV) based Stereo Disparity Pipeline

SAMSUNG
© Samsung 2018

Cost computation

- Pixels dis-similarity
- Lower matching cost, more likely the match

Cost aggregation

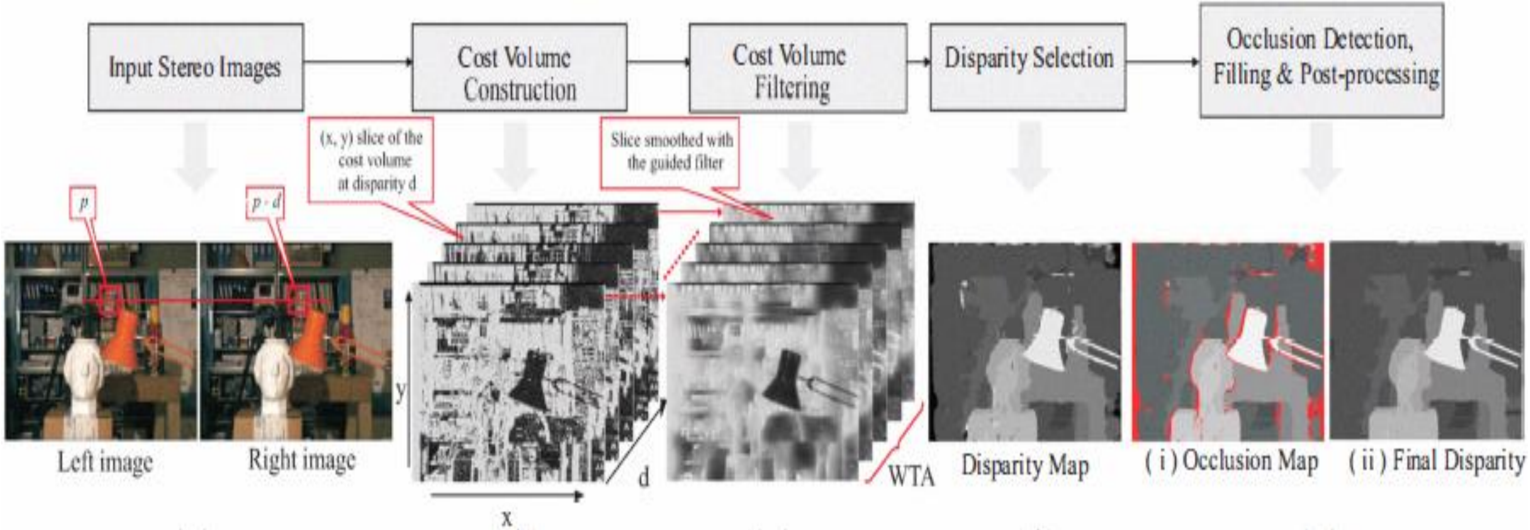
- Find suitable neighborhood
- Aggregation can be weighted

Optimization

- Strategy to decide disparity based on aggregated cost

Post Processing

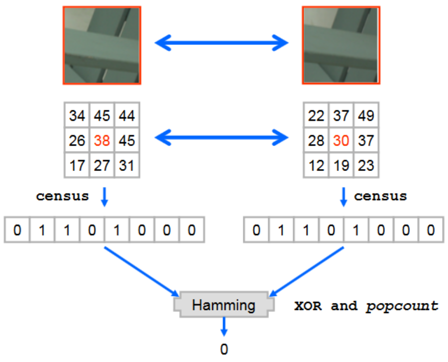
- Handle occlusions
- Smooth and dense depth within objects

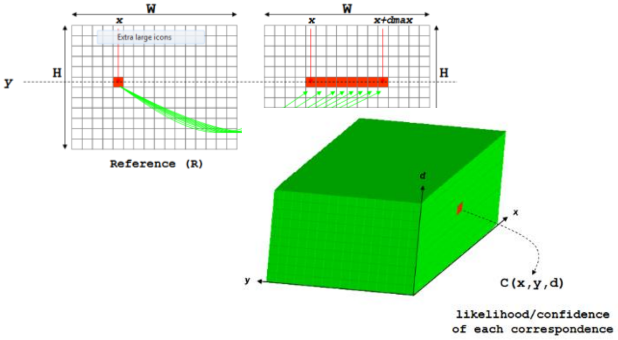


CV based Stereo Disparity – Simple Algorithm

SAMSUNG
© Samsung 2018

Cost computation → Cost aggregation → Optimization → Post Processing





CV based Stereo Disparity – Simple Algorithm

SAMSUNG
© Samsung 2018

Cost computation → Cost aggregation → Optimization → Post Processing

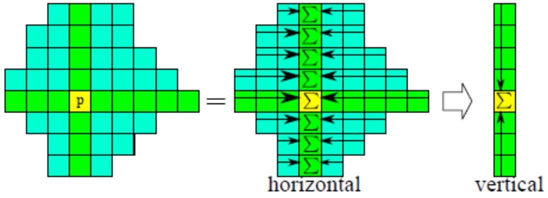
- SAD
- Census transform
- Feature based :SIFT

CV based Stereo Disparity – Simple Algorithm

SAMSUNG
© Samsung 2018

Cost computation → Cost aggregation → Optimization → Post Processing

- SAD
- Census transform
- Feature based :SIFT



CV based Stereo Disparity – Simple Algorithm

SAMSUNG
© Samsung 2018

Cost computation → **Cost aggregation** → **Optimization** → **Post Processing**

Cost computation
- SAD
- Census transform
- Feature based :SIFT

Cost aggregation
- Average over local region
- Cross-arm
- Guided/Bilateral filter

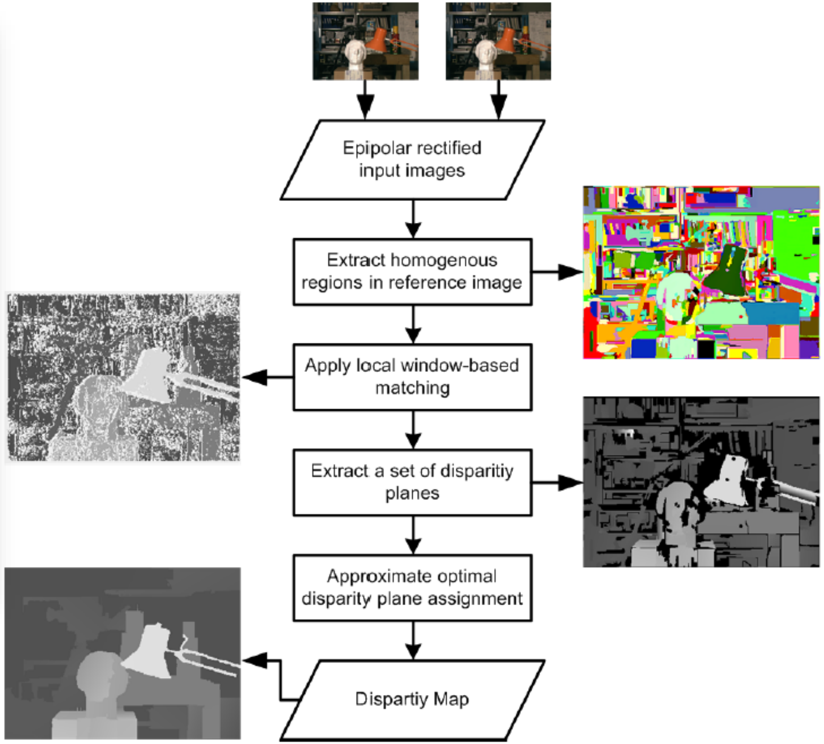
CV based Stereo Disparity – Simple Algorithm

SAMSUNG
© Samsung 2018

Cost computation → **Cost aggregation** → **Optimization** → **Post Processing**

- SAD
- Census transform
- Feature based :SIFT

- Average over local region
- Cross-arm
- Guided/Bilateral filter



CV based Stereo Disparity – Simple Algorithm

© Samsung 2018

Cost computation

- SAD
- Census transform
- Feature based :SIFT

Cost aggregation

- Average over local region
- Cross-arm
- Guided/Bilateral filter

Optimization

- Semi-global matching
- Graph cut
- Belief propagation

Post Processing

CV based Stereo Disparity – Simple Algorithm

SAMSUNG
© Samsung 2018

Cost computation → **Cost aggregation** → **Optimization** → **Post Processing**

Cost computation
- SAD
- Census transform
- Feature based :SIFT

Cost aggregation
- Average over local region
- Cross-arm
- Guided/Bilateral filter

Optimization
- Semi-global matching
- Graph cut
- Belief propagation

Post Processing



CV based Stereo Disparity – Simple Algorithm

SAMSUNG
© Samsung 2018

Cost computation

- SAD
- Census transform
- Feature based :SIFT

Cost aggregation

- Average over local region
- Cross-arm
- Guided/Bilateral filter

Optimization

- Semi-global matching
- Graph cut
- Belief propagation

Post Processing

- L-R consistency
- Subpixel refinement
- Segmentation techniques

Problems in CV Based Approach

SAMSUNG
© Samsung 2018

- ❑ Feature Selection : CENSUS vs SAD vs SIFT vs …
- ❑ Local or Small Neighbourhood Information
- ❑ ONLY Pixel Level Properties → NO SEMANTIC

Evolution → Towards Learning

SAMSUNG
© Samsung 2018

CNN End-To-End : 2016+ — End-To-End system removing separate pre & post processing

- "A Large Dataset to Train Convolutional Networks for Disparity, Optical Flow and Scene Flow Estimation", CVPR 2016

CNN Cost Function : ~2015 — From hand crafted features to learned features, learning similarity between patches

- "A deep visual correspondence embedding model for stereo matching costs", ICCV 2015

Graph Based Methods : ~2011 — Better correspondence searching, enhanced smoothness and occlusion handling

- "Kolmogorov and Zabih's graph cuts stereo matching algorithm", IPOL 2014
  "Pmbp: Patch match belief propagation for correspondence field estimation", IJCV 2014

Semi-global Matching : ~2008 — Smoothness constraint

- "Stereo Processing by Semi-Global Matching and Mutual Information", TPAMI 2008

DEEP LEARNING APPROACHES

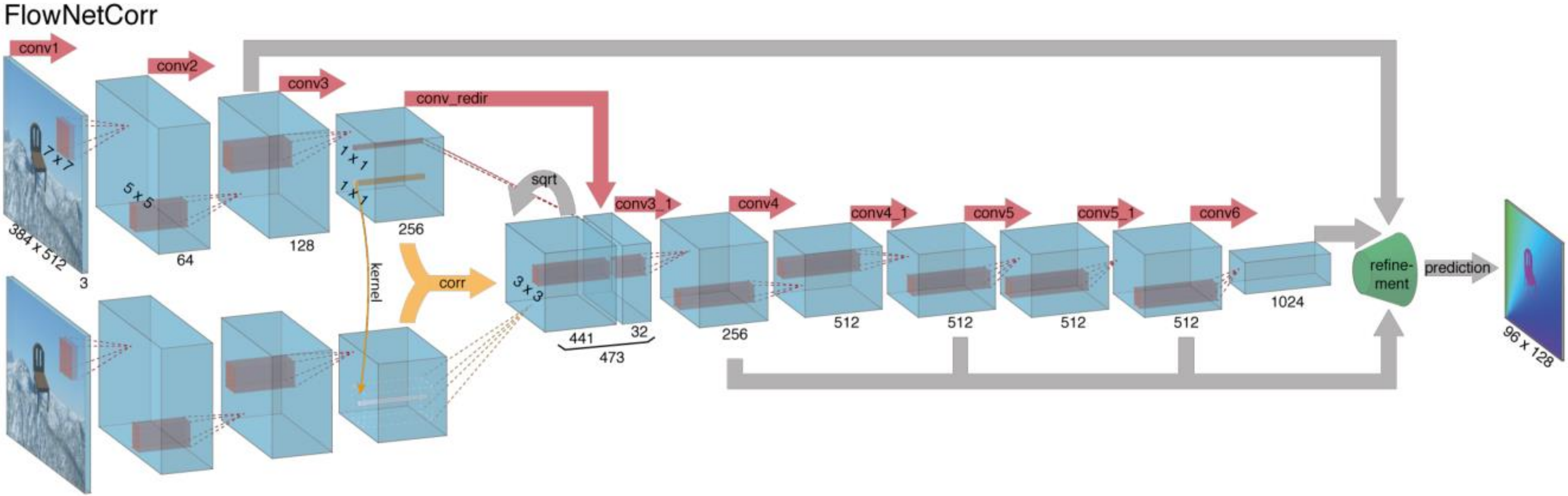
© Samsung 2018

Deep Learning based Stereo Disparity

SAMSUNG
© Samsung 2018

DispNet with Correlation Layer [DispNetC]

- First work: **Mayer, N. et al. A Large Dataset to Train Convolutional Networks for Disparity, Optical Flow and Scene Flow Estimation. CVPR 2016**

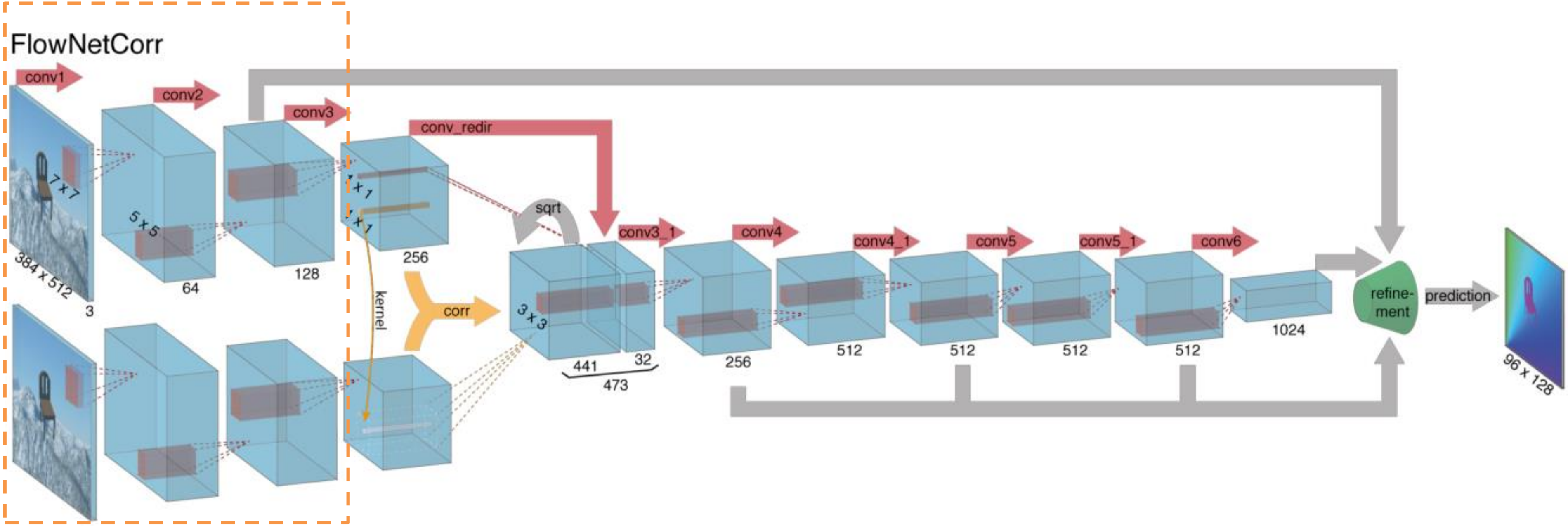


Deep Learning based Stereo Disparity

© Samsung 2018

DispNet with Correlation Layer [DispNetC]

- First work: **Mayer, N. et al. A Large Dataset to Train Convolutional Networks for Disparity, Optical Flow and Scene Flow Estimation. CVPR 2016**



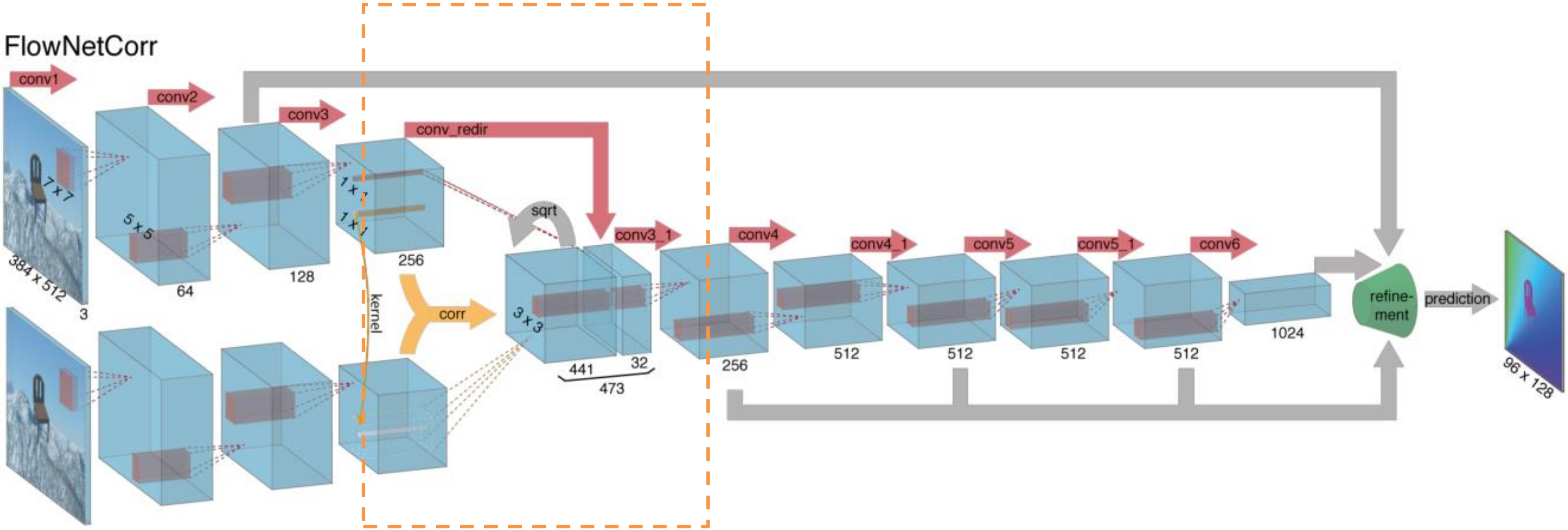
Feature Extraction ➔ CENSUS

Deep Learning based Stereo Disparity

SAMSUNG
© Samsung 2018

DispNet with Correlation Layer [DispNetC]

- First work: **Mayer, N. et al. A Large Dataset to Train Convolutional Networks for Disparity, Optical Flow and Scene Flow Estimation. CVPR 2016**



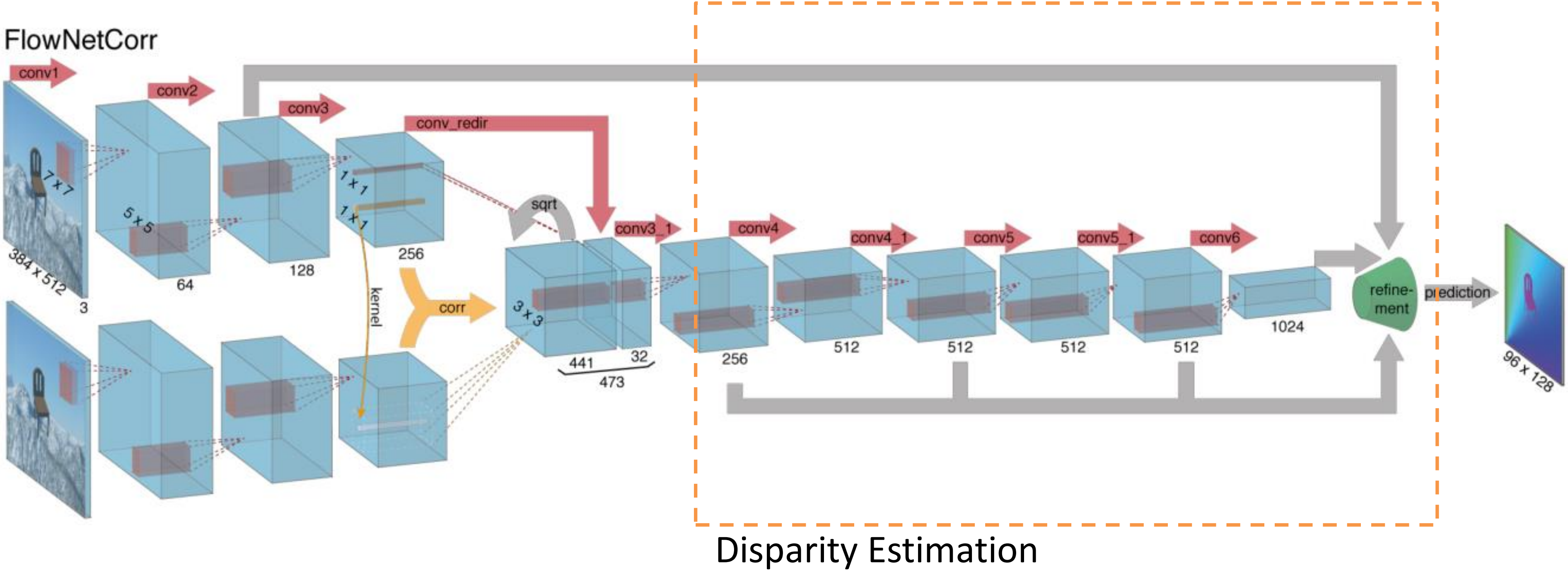
Co-relation Layer ➔ Cost Volume Build

Deep Learning based Stereo Disparity

SAMSUNG
© Samsung 2018

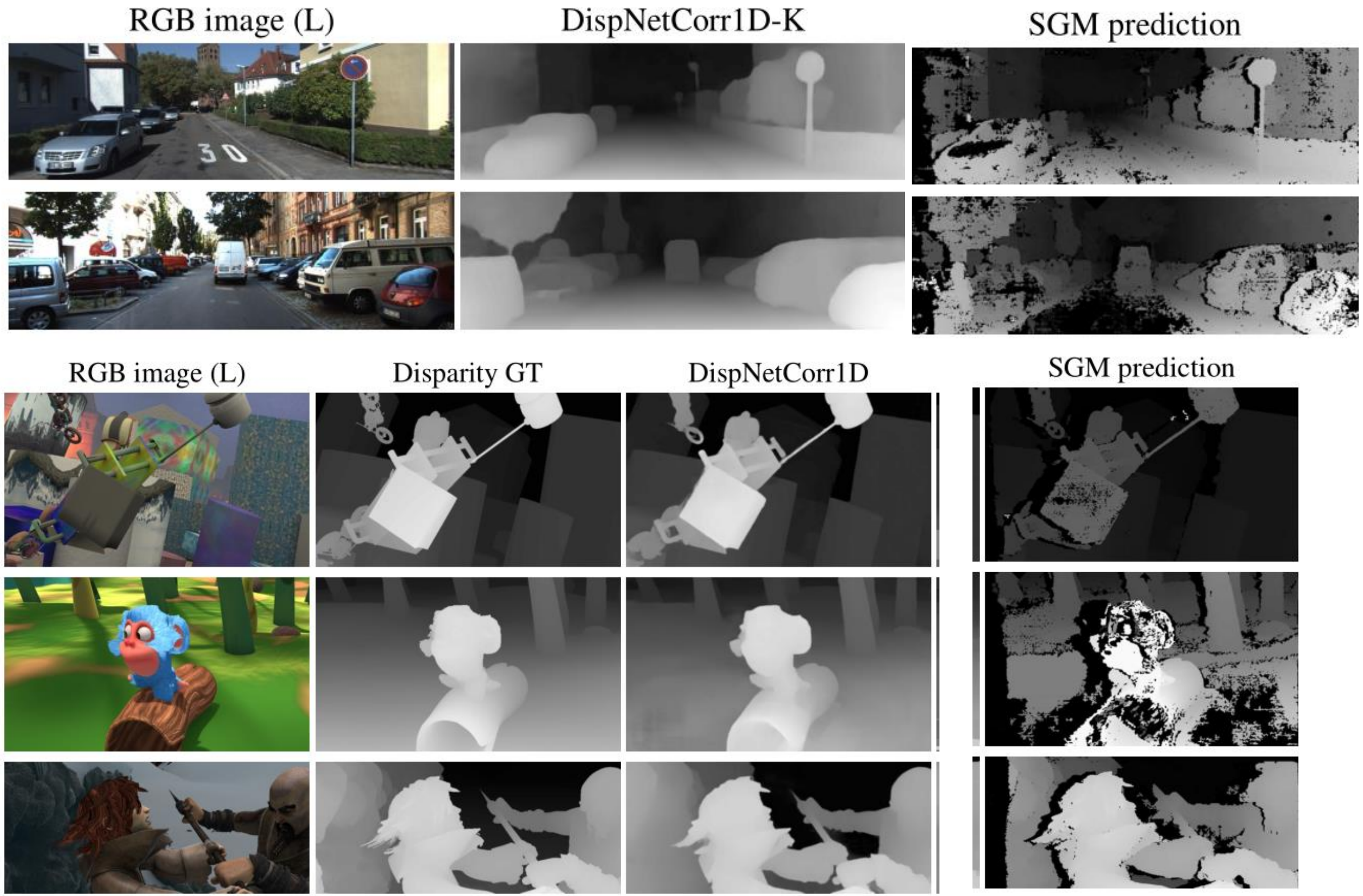
DispNet with Correlation Layer [DispNetC]

- First work: **Mayer, N. et al. A Large Dataset to Train Convolutional Networks for Disparity, Optical Flow and Scene Flow Estimation. CVPR 2016**



Deep Learning : DispNetC Results

© Samsung 2018

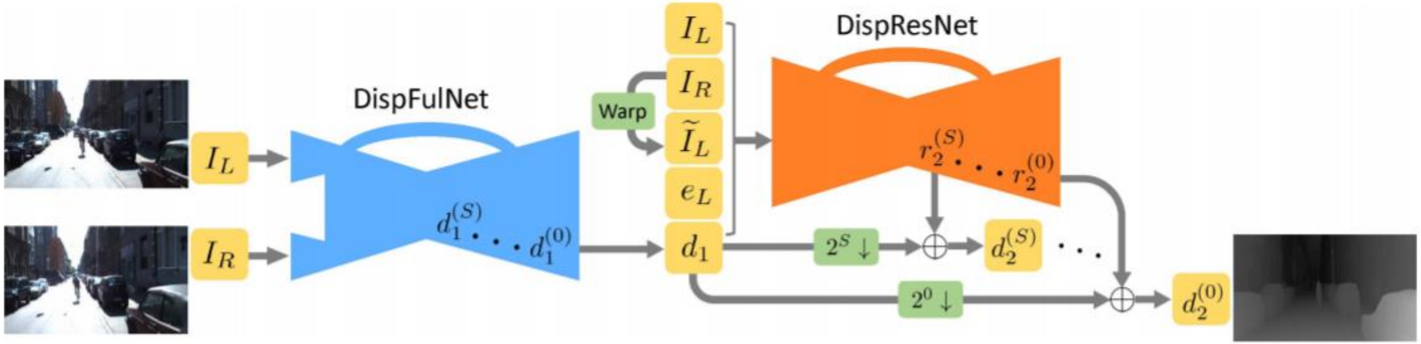


Recent CNN Architectures

SAMSUNG
© Samsung 2018

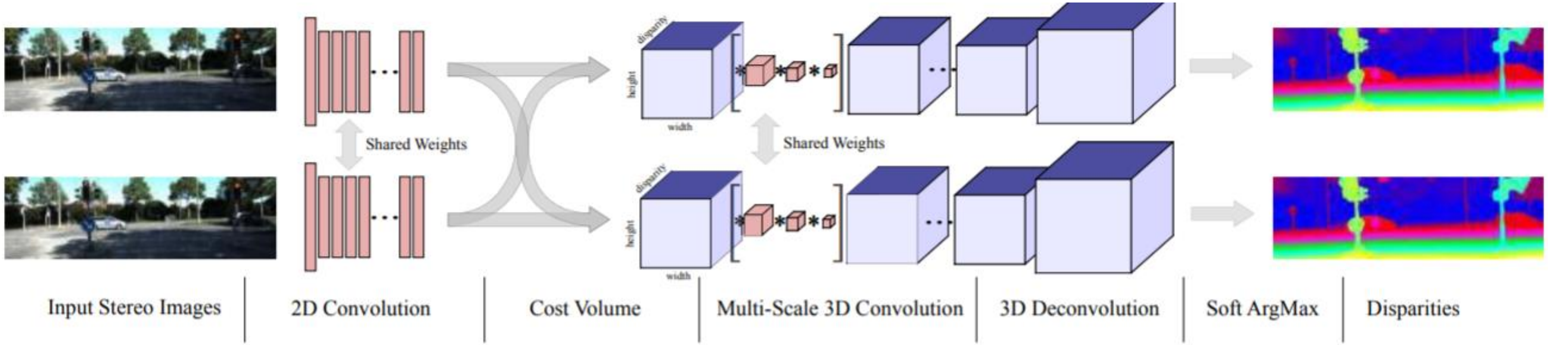
CRL : Cascaded Residual Learning (ICCV 2017)

https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8265317



GC-Net : Geometry and Context Network (ICCV 2017)

https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8237279

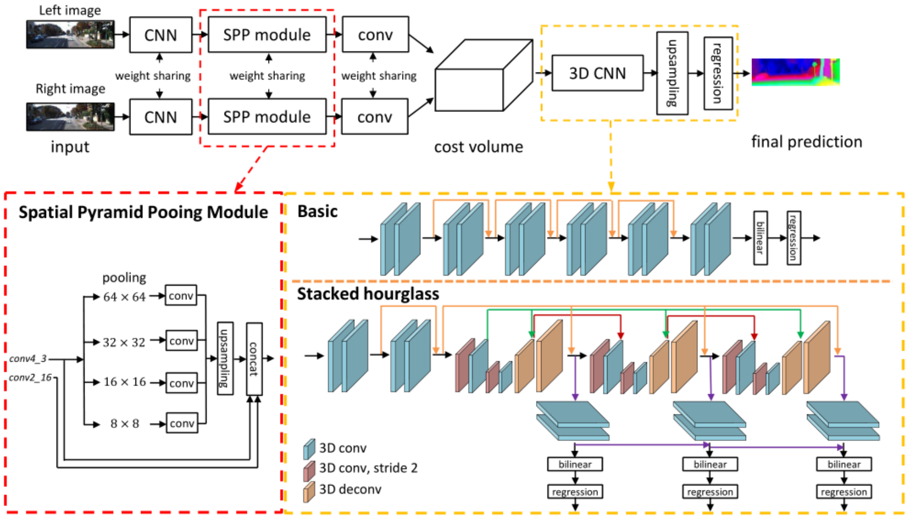


Recent CNN Architectures

SAMSUNG
© Samsung 2018

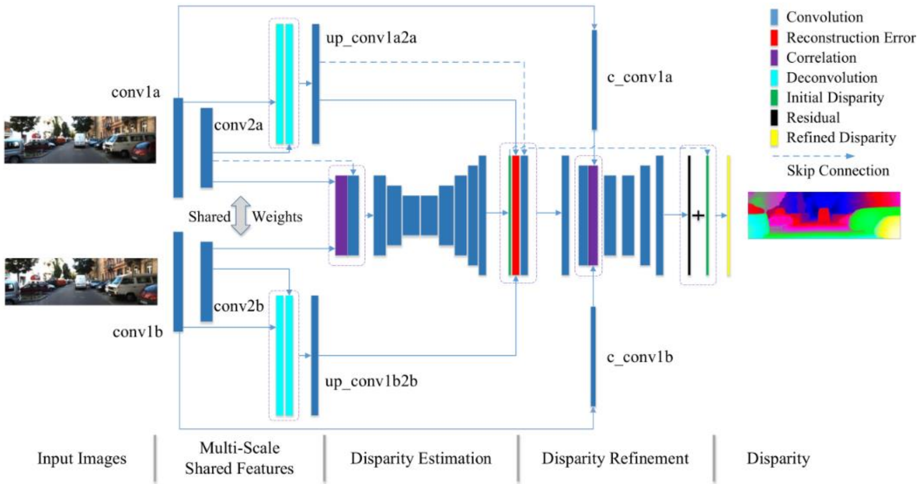
PSM-Net : Pyramid Stereo Matching Network (CVPR 2018)

https://arxiv.org/pdf/1803.08669.pdf



iResNet : Learning for Disparity Estimation through Feature Constancy (CVPR 2018)

https://arxiv.org/pdf/1712.01039.pdf



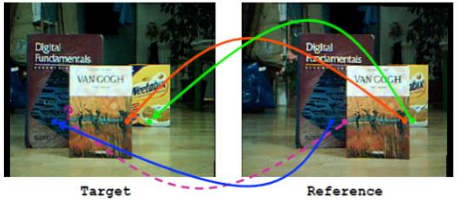
Open Challenges

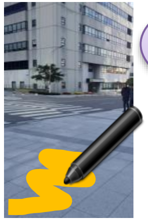
SAMSUNG
© Samsung 2018

Robust Vision Challenge (CVPR 2018) : “ *foster the development of **vision systems** that are **robust** and consequently perform **well** on a **variety of datasets** with different characteristics*”
http://www.robustvision.net/

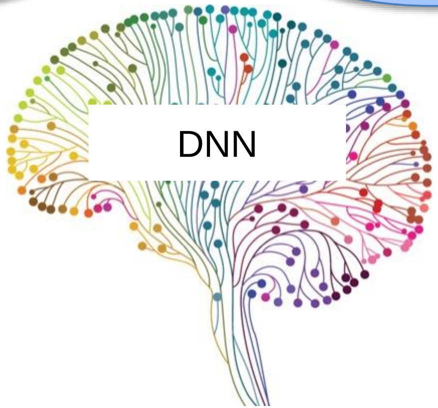
Geometric Corrections

Occlusions





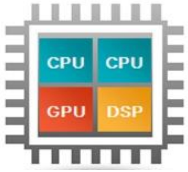
Dense Data Annotation



Specular , Transparent

Uniform Regions

Real Time Performance



SRIB Achievements in Deep Learning based Stereo Depth

© Samsung 2018

Leader Board Position

Low-res two-view results - ETH3D

https://www.eth3d.net/low_res_two_view?mask=all&metric=bad-4-0

ETH3D Benchmark Home Documentation Datasets Benchmarks About Submit

Coverage: dense **Set**: Test **Metric**: 99% quantile [px] **Mask**: all Download this table as CSV

Method	all	lakes. 1l	lakes. 1s	sand box 1l	sand box 1s	stora. room 1l	stora. room 1s	stora. room 2l	stora. room 2s	stora. room 2 1l	stora. room 2 1s	stora. room 2 2l	stora. room 2 2s	stora. room 3l	stora. room 3s	tunnel 1l	tunnel 1s	tunnel 2l
DISCO	**2.40** 1	0.99 4	11.99 2	1.81 22	1.12 16	3.47 5	1.24 6	**3.34** 1	2.57 3	2.73 7	1.96 5	4.52 20	1.90 10	4.18 5	3.62 9	0.37 2	0.27 4	0.31 3
DLCB_ROB	2.55 2	1.32 9	**10.26** 1	1.31 7	1.04 14	**2.94** 1	1.29 7	3.81 4	2.73 4	2.93 10	2.17 9	3.37 11	1.53 5	10.40 15	3.26 7	0.42 5	0.35 7	0.41 7
iResNet_ROB	2.72 3	10.38 29	14.41 3	1.46 13	0.99 9	3.74 12	0.66 2	4.01 6	2.84 6	2.04 2	2.05 7	1.98 2	1.38 3	4.73 6	1.91 3	**0.35** 1	0.20 2	0.28 2
LALA_ROB	2.88 4	2.13 15	19.08 12	**1.10** 1	0.99 9	3.91 17	1.81 15	4.50 20	3.04 10	3.32 14	1.87 3	2.75 7	1.66 6	4.06 3	4.02 11	0.47 10	0.44 15	0.55 20
DN-CSS_ROB	2.89 5	9.72 26	19.71 15	2.07 25	0.90 6	4.01 19	**0.61** 1	4.06 7	3.35 15	**1.66** 1	1.89 4	**1.39** 1	**1.06** 1	**3.24** 1	1.88 2	0.44 7	**0.19** 1	0.44 9
ETE_ROB	2.93 6	2.12 14	17.80 9	1.13 2	0.94 8	3.71 11	1.53 10	4.29 12	2.98 8	3.32 14	2.13 8	2.34 4	1.51 4	8.68 9	2.82 6	0.44 7	0.40 11	0.51 17
NCCL2	2.96	4.07	15.75	1.24	1.03	3.48	2.55	3.79	**2.39**	2.78	1.80	2.41	1.67	10.27	2.48	0.46	0.41	0.57

vision.middlebury.edu/stereo/eval3/

Not secure | vision.middlebury.edu/stereo/eval3/

Stereo Evaluation • Datasets • Code • Submit

Middlebury Stereo Evaluation - Version 3

Mouseover the table cells to see the produced disparity map. Clicking a cell will blink the ground truth for comparison. To change the table type, click the links below. For more information, please see the description of new features.

Submit and evaluate your own results. See snapshots of previous results. See the evaluation v.2 (no longer active).

Set: test dense test sparse training dense training sparse
Metric: bad 0.5 bad 1.0 bad 2.0 bad 4.0 avgerr rms A50 A90 A95 A99 time time/MP time/GD
Mask: nonocc all

plot selected show invalid **Reset sort** **Reference list**

A99 (pixels)

Date	Name	Res	Weight Avg	Austr MP: 5.6 nd: 290 im0 im1 GT nonocc	AustrP MP: 5.6 nd: 290 im0 im1 GT nonocc	Bicyc2 MP: 5.6 nd: 250 im0 im1 GT nonocc	Class MP: 5.7 nd: 610 im0 im1 GT nonocc	ClassE MP: 5.7 nd: 610 im0 im1 GT nonocc	Compu MP: 1.5 nd: 256 im0 im1 GT nonocc	Crusa MP: 5.5 nd: 800 im0 im1 GT nonocc	CrusaP MP: 5.5 nd: 800 im0 im1 GT nonocc	Djemb MP: 5.7 nd: 320 im0 im1 GT nonocc	DjembL MP: 5.7 nd: 320 im0 im1 GT nonocc	Hoops MP: 5.7 nd: 410 im0 im1 GT nonocc	Livgrm MP: 5.9 nd: 320 im0 im1 GT nonocc	Nkuba MP: 5.5 nd: 570 im0 im1 GT nonocc	Plants MP: 5.6 nd: 320 im0 im1 GT nonocc	Stairs MP: 5.2 nd: 450 im0 im1 GT nonocc
05/22/18	DN-CSS_ROB	H	**82.0** 1	118 2	**107** 1	**40.5** 1	71.4 4	146 9	**29.0** 1	99.9 2	105 3	21.6 2	66.4 18	171 30	57.3 8	65.7 2	**113** 1	127 7
10/10/18	DISCO	H	86.6 2	**108** 1	110 2	46.1 2	102 7	146 8	31.0 2	108 3	101 2	**17.1** 1	104 30	166 18	**51.5** 1	68.6 3	142 3	87.2 3
05/31/18	iResNet_ROB	H	87.5 3	121 4	118 3	55.7 3	86.6 5	114 2	95.0 20	**70.9** 1	**73.7** 1	32.1 10	47.5 10	172 31	59.2 12	**62.1** 1	142 2	145 8
05/26/18	NOSS_ROB	H	104 4	130 21	125 8	71.2 8	**63.5** 1	137 5	97.0 25	126 9	123 6	41.7 24	41.9 3	152 9	58.1 10	106 13	177 29	174 26
03/06/18	NOSS	H	104 5	130 21	125 8	71.2 8	**63.5** 1	137 5	97.0 25	126 9	123 6	41.7 24	41.9 3	152 9	58.1 10	106 13	177 29	174 27
05/01/18	PSMNet_ROB	Q	106 6	145 63	143 64	84.1 27	106 8	128 3	58.0 5	112 4	111 4	34.3 12	139 43	169 24	62.5 13	98.4 6	163 9	118 6
12/11/17	OVOD	H	108 7	120 3	118 4	69.4 4	122 10	183 16	99.0 35	129 15	130 14	34.8 14	55.3 14	149 3	64.5 19	80.5 4	168 14	153 9
06/22/17	LocalExp	H	109 8	128 16	126 12	70.9 7	89.5 6	175 13	97.0 25	134 19	125 8	37.6 19	43.8 5	158 15	63.9 18	108 15	175 21	163 15

**As on Oct-2018

References

© Samsung 2018

More refernces on Stereo :CV method

- H. Hirschmueller. Stereo processing by semi-global matching and mutual information. PAMI 30(2):328-341, 2008
- S. Drouyer, et al. Sparse stereo disparity map densifycation using hierarchical image segmentation. 13th International Symposium on Mathematical Morphology.
- L. Li, X. Yu, S. Zhang, X. Zhao, and L. Zhang. 3D cost aggregation with multiple minimum spanning trees for stereo matching. Applied Optics 56(12):3411-3420, 2017.
- L. Li, S. Zhang, X. Yu, and L. Zhang. PMSC: PatchMatch-based superpixel cut for accurate stereo matching. IEEE Trans on Circuits and Systems for Video Technology, 2016.
- "Multiview Geometry in Computer Vision", book by Hartley and Zisserman

More refernces on Stereo : End to end CNN method

- J. Chang and Y. Chen: Pyramid Stereo Matching Network. arXiv preprint arXiv:1803.08669 2018.
- Z. Liang, Y. Feng, Y. Guo and H. Liu: Learning for Disparity Estimation through Feature Constancy. arXiv preprint arXiv:1712.01039 2017.
- J. Pang, et al: Cascade residual learning: A two-stage convolutional neural network for stereo matching. ICCV Workshop on Geometry Meets Deep Learning 2017.

Tutorial Material on Stereo

- http://www.cse.psu.edu/~rtc12/CSE486/lecture09.pdf
- http://www.inf.u-szeged.hu/~kato/teaching/computervision/02-CameraGeometry.pdf
- http://www.ics.uci.edu/~majumder/vispercep/chap8notes.pdf
- http://vision.deis.unibo.it/~smatt/Seminars/StereoVision.pdf
- https://courses.cs.washington.edu/courses/cse455/09wi/Lects/lect16.pdf

Stereo Datasets:

- Middlebury: http://vision.middlebury.edu/stereo/eval3/
- Kitti: http://www.cvlibs.net/datasets/kitti/eval_scene_flow.php?benchmark=stereo
- SceneFlow: https://lmb.informatik.uni-freiburg.de/resources/datasets/SceneFlowDatasets.en.html
- ETH dataset: https://www.eth3d.net/low_res_two_view

Acknowledgements

© Samsung 2018

- Extend sincere gratitude to following members for providing their valuable support and help
  - Kunal Swami
  - Dr. Rituparna Sarkar
  - Yash Harbhajanka
  - Bhushan Bhagwan Gawde
  - Dr. Lokesh Boregowda

- Images are borrowed from various sources and internet
  - *"Stereo Vision: Algorithms and Applications" by Stefano Mattocia*
  - *"On Building an Accurate Stereo Matching System on Graphics Hardware" by Xing Mei*
  - *Etc…*


THANK
YOU