EFFECT OF SPARSITY ON SCENE FLOW ESTIMATION

Chandra Teja Kommineni

Sai Rahul Vaddadi

Tasks and Contributions

TASKS	CONTRIBUTION
Literature survey of architectures	Rahul / Chandra
Datasets preparation & pre-processing	Chandra
Models training & Validation	Rahul
Results & Inference	Rahul
Slides preparation	Chandra
Final report	Rahul / Chandra

Objectives

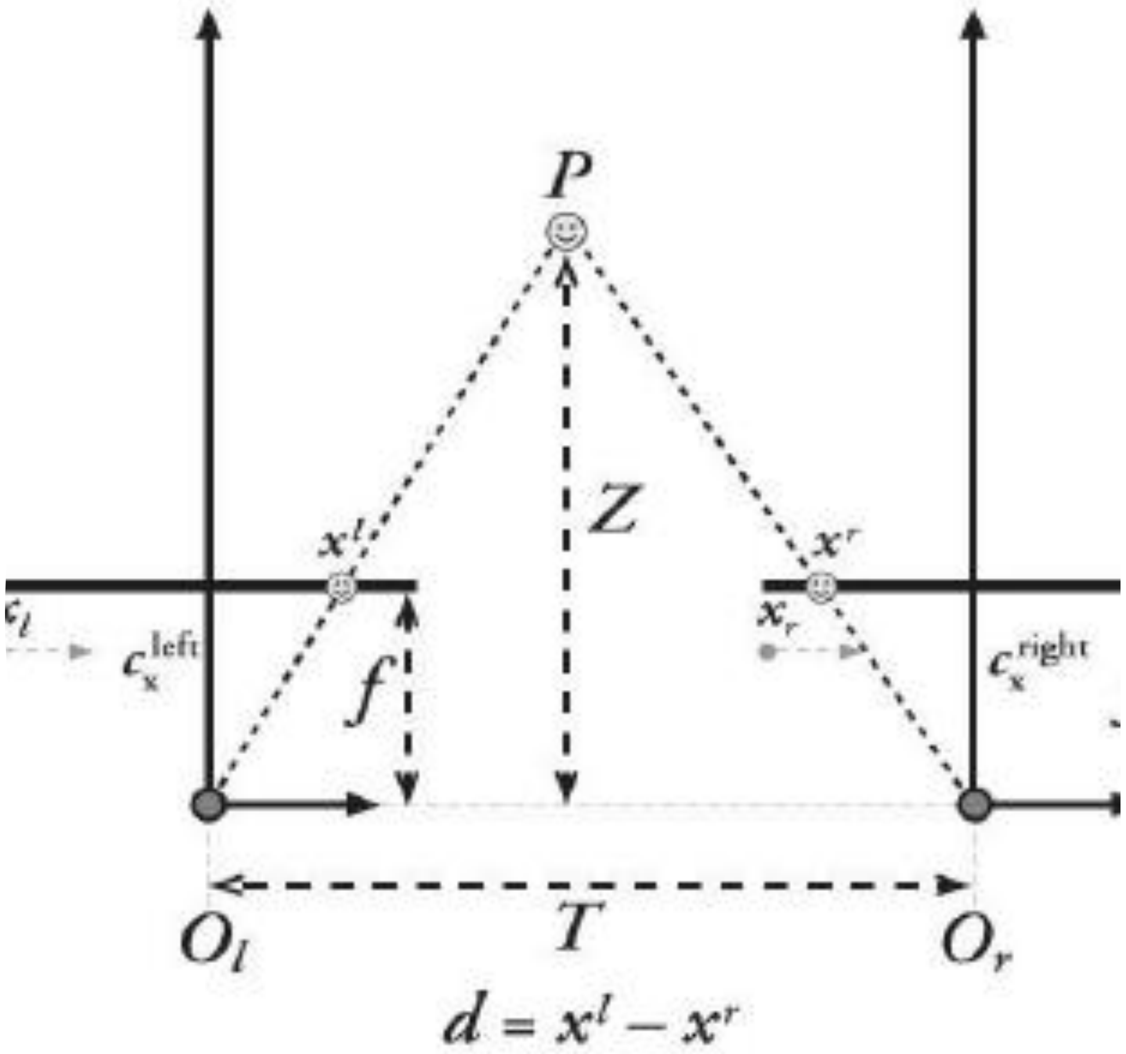
- Estimating scene flow by independent estimation of optical flow, disparities at reference frame & next time step.
- Evaluate the performance using Endpoint Error (EPE) as Metric
- Analyze the effect of sparsity in input images on scene flow estimation.

What is Optical Flow ?

- Optical flow is a projection of the world's 3D motion onto the image plane.
- Constraint equation for optical flow $\mathbf{I_x u + I_y v + I_t = 0}$, where u and v are the x and y components of the optical flow vector

What is disparity ?

Disparity refers to the difference in coordinates of similar features within two stereo images.



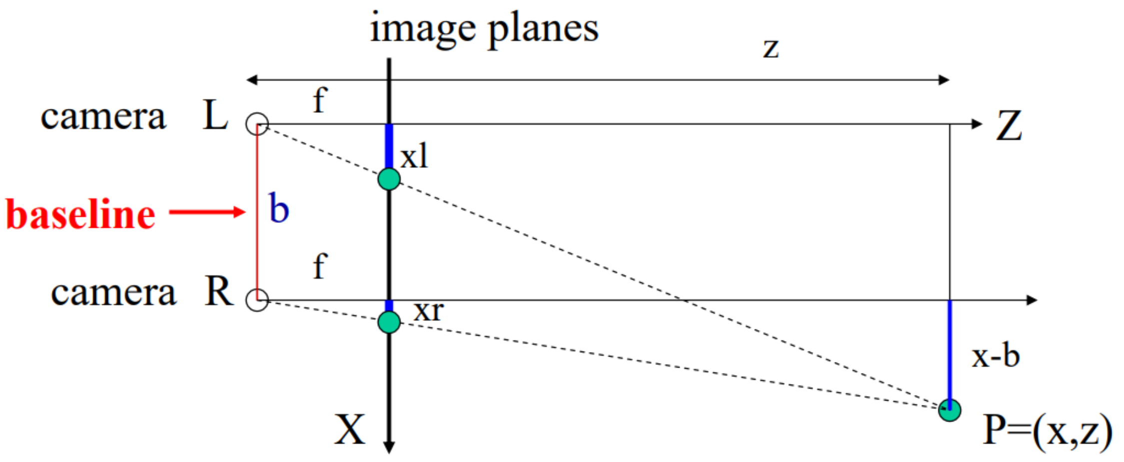
Depth from disparity:

- Depth can be estimated from disparity d by method of triangulation

Depth $z = f*b / (xl - xr) = f*b/d$

$x = xl*z/f$ or $b + xr*z/f$

$y = yl*z/f$ or $yr*z/f$

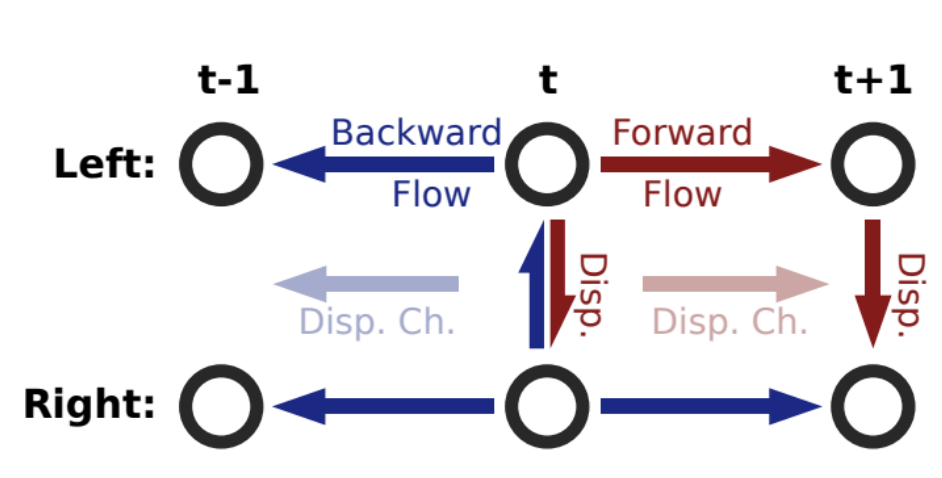


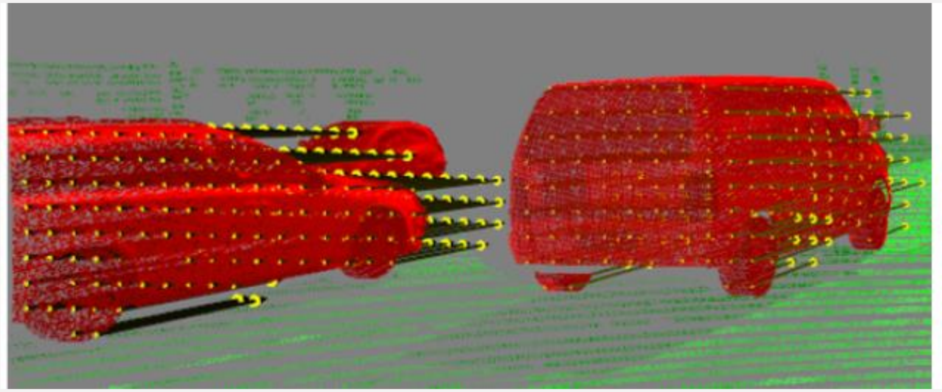
What is scene flow?

- Scene flow is underlying 3D motion field that can be computed from stereo videos or RGBD videos.
- Estimating scene flow means providing the depth and 3D motion vectors of all visible points in a stereo video. In this project we'll be focusing on forward flow & disparities.

Thus, this task splits into

- Optical flow
- Disparity at reference frame
- Disparity change (fills gaps left by occlusions)



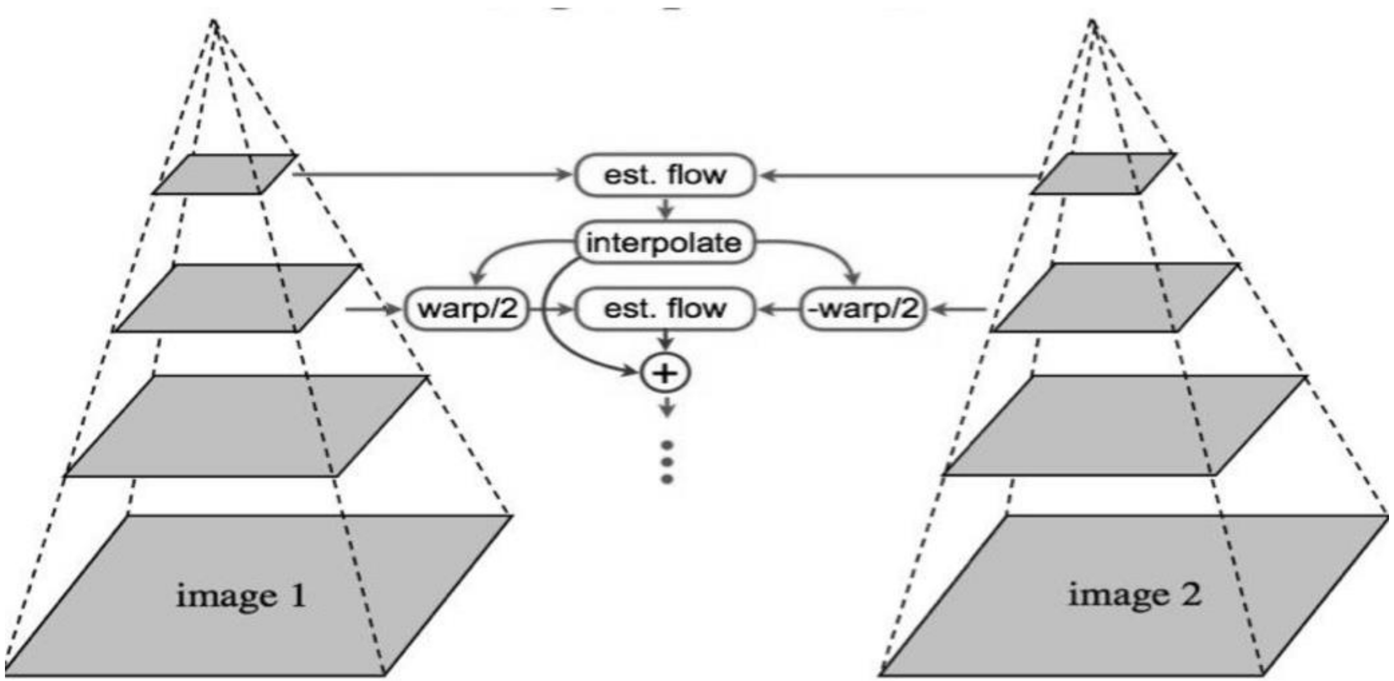


Traditional methods

- Energy Minimization - Horn & Schunck (Computationallly expensive for realtime applications)
- Lucas-Kanade Methods – Bruce D.Lucas & Kanade (sparse optical flow)

CNN based Methods (FlowNetS, FlowNet Corr) :

- FlowNetS uses architecture similar to U-Net architecture on raw Images
- FlownetS performs better for large displacements in optical flow



Coarse-to-fine flow estimation

Source: Survey of local and global optical flow with coarse to fine method paper

Dataset:

- We cannot obtain ground truth for real world, since it is not possible to manually label optical flow.

- To compensate for the lack of data we work with synthetic data.

- Ground truth can be obtained in simulated data, since location of objects and transformation are known.

- Data Augmentation: To reduce data scarcity and prevent overfitting, we can perform translation, rotation, scaling, changing brightness & colour.

- Driving dataset:

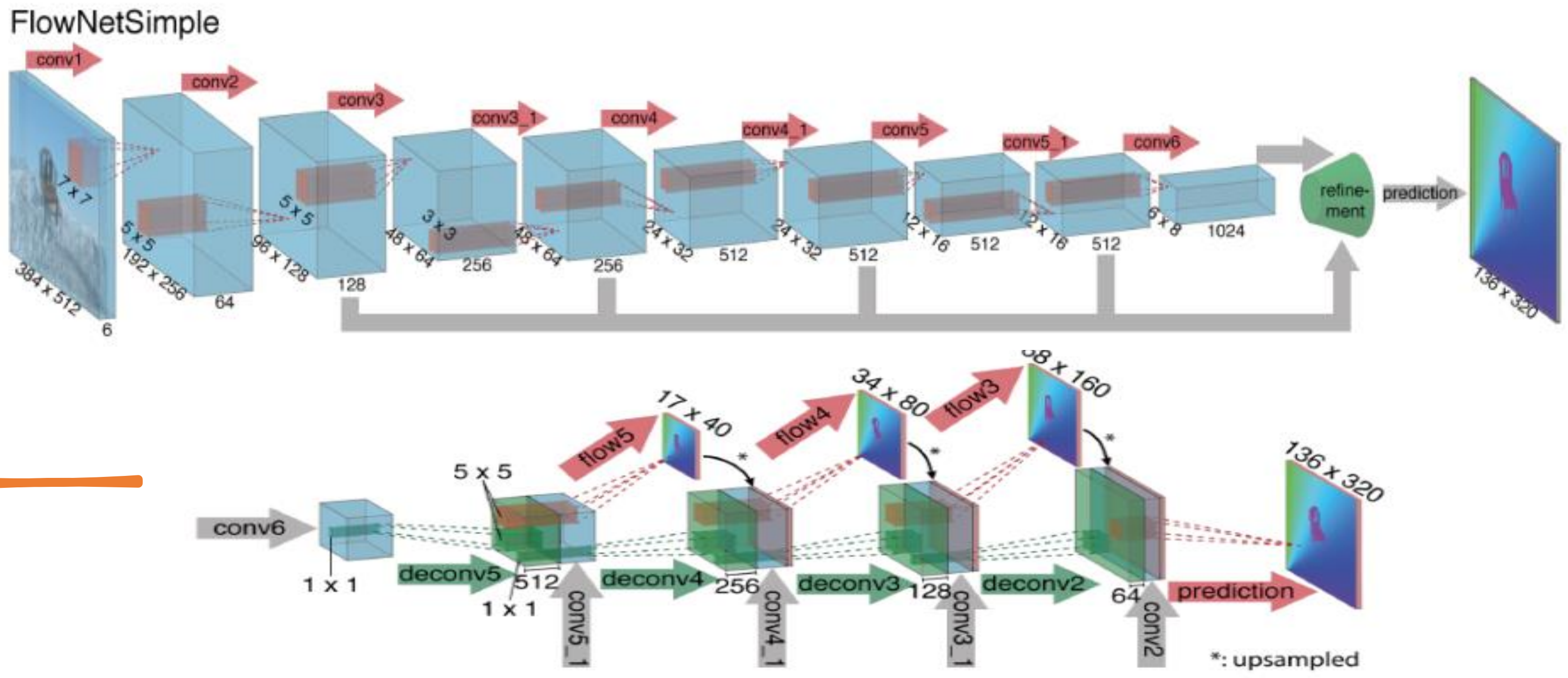
- The Driving scene is a mostly naturalistic, dynamic street scene from
- the viewpoint of a driving car, made to resemble the KITTI datasets.

TTI 2015 Driving (o

https://lmb.informatik.uni-freiburg.de/resources/datasets/SceneFlowDatasets.en.html

Flow net architecture

- Left images at t0 and t1 are stacked as input
- Contracting network: feature representation and matching info are learnt from input images
- Refinement Network - features are scaled into per pixel optical flow prediction
- Unpooling layer extends the feature maps and increases the resolution



Datasets	Driving
Evaluation Metrices	End Point Error
Optimizer	Adam
Learning Rate	1e-4 (default)
Batch Size	1
Initial training	300 images for 10 epochs

DispNet architecture

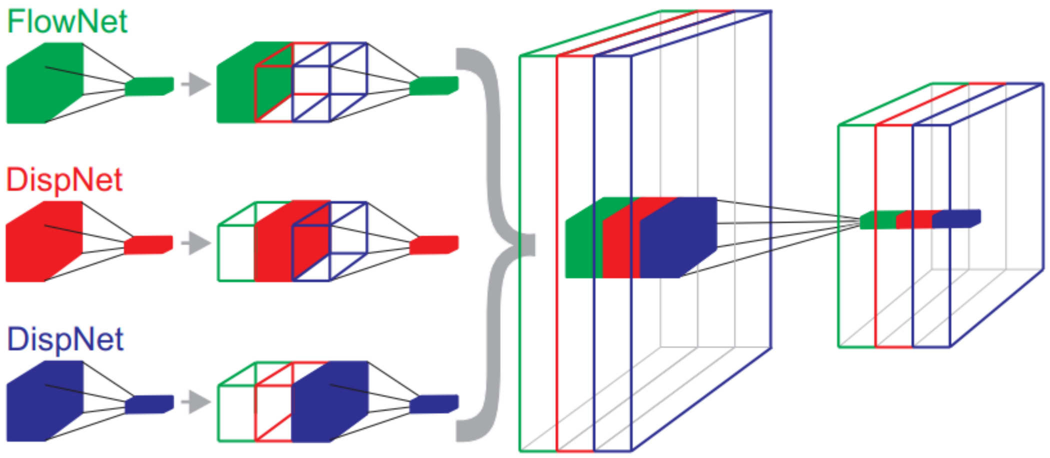
Name	Kernel	Str.	Ch I/O	InpRes	OutRes	Input
conv1	7×7	2	$6/64$	768×384	384×192	images
conv2	5×5	2	$64/128$	384×192	192×96	conv1
conv3a	5×5	2	$128/256$	192×96	96×48	conv2
conv3b	3×3	1	$256/256$	96×48	96×48	conv3a
conv4a	3×3	2	$256/512$	96×48	48×24	conv3b
conv4b	3×3	1	$512/512$	48×24	48×24	conv4a
conv5a	3×3	2	$512/512$	48×24	24×12	conv4b
conv5b	3×3	1	$512/512$	24×12	24×12	conv5a
conv6a	3×3	2	$512/1024$	24×12	12×6	conv5b
conv6b	3×3	1	$1024/1024$	12×6	12×6	conv6a
pr6+loss6	3×3	1	$1024/1$	12×6	12×6	conv6b
upconv5	4×4	2	$1024/512$	12×6	24×12	conv6b
iconv5	3×3	1	$1025/512$	24×12	24×12	upconv5+pr6+conv5b
pr5+loss5	3×3	1	$512/1$	24×12	24×12	iconv5
upconv4	4×4	2	$512/256$	24×12	48×24	iconv5
iconv4	3×3	1	$769/256$	48×24	48×24	upconv4+pr5+conv4b
pr4+loss4	3×3	1	$256/1$	48×24	48×24	iconv4
upconv3	4×4	2	$256/128$	48×24	96×48	iconv4
iconv3	3×3	1	$385/128$	96×48	96×48	upconv3+pr4+conv3b
pr3+loss3	3×3	1	$128/1$	96×48	96×48	iconv3
upconv2	4×4	2	$128/64$	96×48	192×96	iconv3
iconv2	3×3	1	$193/64$	192×96	192×96	upconv2+pr3+conv2
pr2+loss2	3×3	1	$64/1$	192×96	192×96	iconv2
upconv1	4×4	2	$64/32$	192×96	384×192	iconv2
iconv1	3×3	1	$97/32$	384×192	384×192	upconv1+pr2+conv1
pr1+loss1	3×3	1	$32/1$	384×192	384×192	iconv1

Datasets	Driving
Evaluation Metrices	End Point Error
Optimizer	Adam
Learning Rate	1e-2
Batch Size	4
Initial training	300 image pairs for 50 epochs

Scene flow architecture

- The Sceneflow was estimated by stacking the independent predictions of optical flow and disparities at t0 and t1 obtained from our trained Flow and Disparity Network architectures respectively.

- The result depicts the sceneflow in the image plane and not the real-world coordinate system.



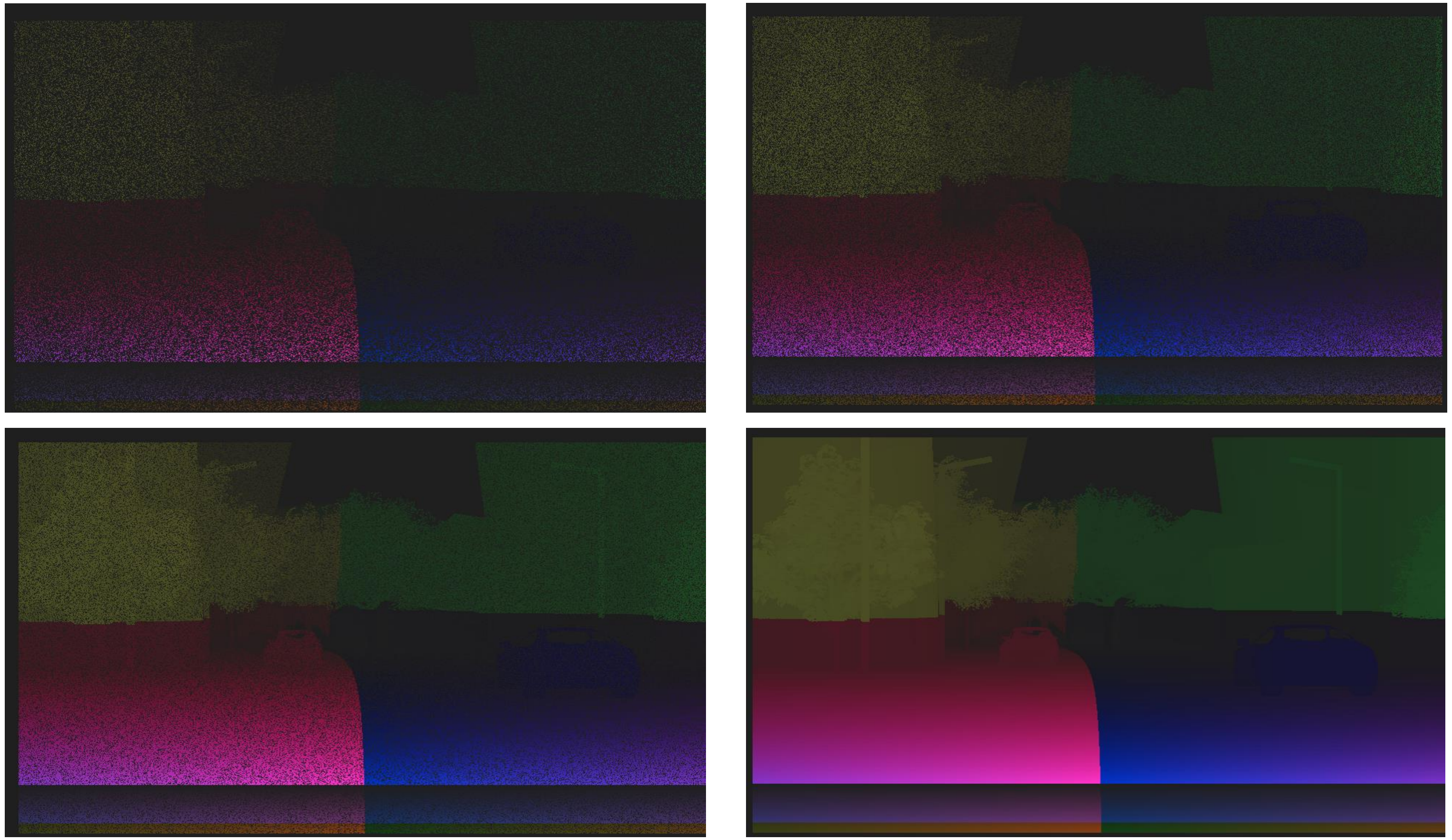
Results

The tables below depict the EPE of our scene flow estimation (on scale of 1e+12) for different sparsity of input images and ground truths, followed by comparison of our optical flow results with the baseline method.

Sparsity	Our Scene flow (EPE)	
	Driving	Kitti2015
25%	107.47	11.97
50%	98.04	3.42
75%	5.2	26.41
100%	2e-7	2.46e-7

Method	Optical Flow (EPE)
Baseline	22.5
Ours	10.5

Expected Scene Flow Results



- These are the expected scene flow results for different 25%, 50%, 75% and 100% dense input and ground truths respectively.

Future work

Train on flying things 3D and validate on sintel dataset (can improve EPE).

PWC Net for computing optical flow:

- 17 times smaller and 2 times faster than FlowNet 2.0
- Easier to train than FlowNet 2.0 & SpyNet.

References

- S. Baker, D. Scharstein, J. Lewis, S. Roth, M. J. Black, and R. Szeliski. A database and evaluation methodology for optical flow. Technical Report MSR-TR-2009-179, December 2009.
- D. J. Butler, J. Wulff, G. B. Stanley, and M. J. Black. A naturalistic open source movie for optical flow evaluation. In ECCV, Part IV, LNCS 7577, pages 611–625, Oct. 2012.
- J. Cech, J. Sanchez-Riera, and R. P. Horaud. Scene flow estimation by growing correspondence seeds. In CVPR, 2011.
- A. Dosovitskiy, P. Fischer, E. Ilg, P. Hausser, C. Hazırbas¸, ¨ V. Golkov, P. van der Smagt, D. Cremers, and T. Brox. FlowNet: Learning optical flow with convolutional networks. In ICCV, 2015.
- A. Dosovitskiy, J. T. Springenberg, and T. Brox. Learning to generate chairs with convolutional neural networks. In CVPR, 2015.
- D. Eigen, C. Puhrsch, and R. Fergus. Depth map prediction
- Nikolaus Mayer, Ed dy Ilg, Philip Häusser, Philipp Fischer, Daniel Cremers, Alexey Dosovitskiy, Thomas Brox . A Large Dataset to Train Convolutional Networks for Disparity, Optical Flow, and Scene Flow Estimation.