Into the Twilight Zone: Depth Estimation using Joint Structure-Stereo Optimization

Aashish Sharma, Loong-Fah Cheong

Department of ECE, National University of Singapore

ECCV 2018 European Conference on Computer Vision

8 – 14 September 2018 | Munich, Germany

Motivation –

Stereo vision under night-time conditions –

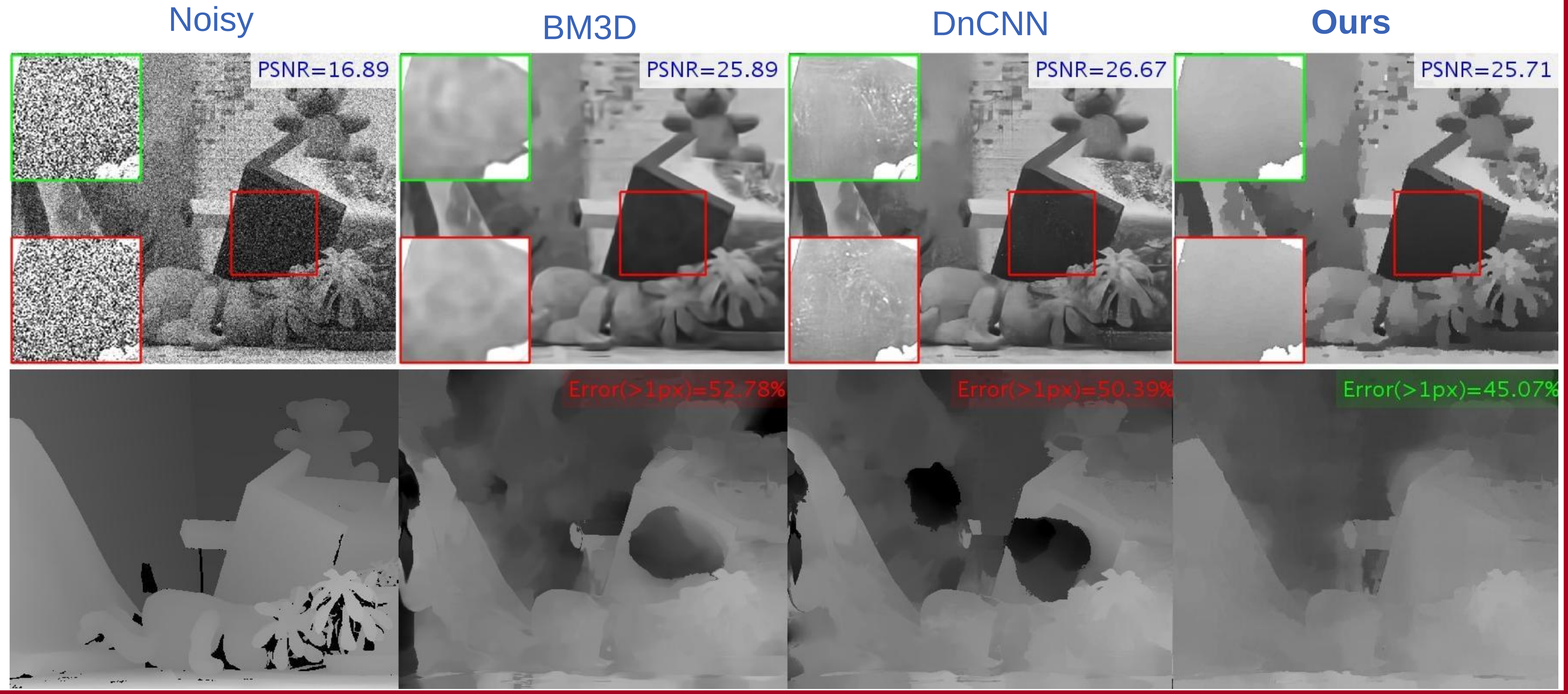
- An important but a relatively unexplored area.
- Manifold challenges – severe noise, varying glow and glare, multiple moving light-sources, lens-flares; no reliable datasets, etc.
- We propose to handle the problem of noise first in this paper.

Problems with conventional denoising –

- PSNR not an optimal measure for envisioning stereo quality.
- Maximizing PSNR for maximizing stereo quality is not necessary. Low-PSNR but more stereo-consistent images can still perform better.

Our proposed solution –

- Obtain piecewise-constant structures from the images by $TV - L_2$ style of decomposition. Give-away the fine-details (and some PSNR), as they are inextricably mixed with noise, and hence, more difficult to recover.
- Obtain the structures stereo consistently by injecting a BCC+GCC based stereo constraint into the above decomposition objective.



Problem Formulation –

- To obtain disparity D_2 using structures I_{s1}, I_{s2} from input images I_{n1}, I_{n2}

$$E_{ALL} = \underbrace{E_{StructureData}(I_{s1}, I_{s2}) + \lambda_S \cdot E_{StructureSmooth}(I_{s1}, I_{s2}) + \lambda_{SD} \cdot E_{StereoData}(I_{s1}, I_{s2}, D_2)}_{E_{Structure}(I_{s1}, I_{s2}, D_2^*)} + \lambda_{SS} \cdot E_{StereoSmooth}(D_2)$$

$$\underbrace{\lambda_{SD} \cdot E_{StereoData}(I_{s1}, I_{s2}, D_2) + \lambda_{SS} \cdot E_{StereoSmooth}(D_2)}_{E_{Stereo}(I_{s1}^*, I_{s2}^*, D_2)}$$

(* is used to denote constant variables)

Optimization –

RTV - Relative Total Variation [6]

$$E_{StructureData}(I_{s1}, I_{s2}) = \sum_p \left((I_{s1}(p) - I_{n1}(p))^2 + (I_{s2}(p) - I_{n2}(p))^2 \right)$$

$$E_{StructureSmooth}(I_{s1}, I_{s2}) = \sum \left(\mathbf{RTV}(I_{s1}(p)) + \mathbf{RTV}(I_{s2}(p)) \right)$$

$TV - L_2$ based decomposition (using **RTV**$(\cdot)$ for robustness)

$$E_{StereoData}(I_{s1}, I_{s2}, D_2^*) = \sum_p \left(\alpha \cdot \sum_{q \in W_p} \left(I_{s2}(q) - I_{s1}\left(q - D_2^*(q)\right) \right)^2 + \sum_{q \in W_p} \min\left(\left| \nabla I_{s2}(q) - \nabla I_{s1}(q - D_2^*(q)) \right|, \theta \right) \right)$$

'BCC+GCC' stereo constraint (also data term for stereo)

$$E_{StereoSmooth}(D_2) = \sum_p \sum_{q \in N4_p} \begin{cases} \lambda_{SS1}, & \text{if } \left[|D_2(p) - D_2(q)| = 1 \right] \\ \lambda_{SS2}, & \text{if } \left[|D_2(p) - D_2(q)| > 1 \right] \end{cases}$$

smoothness term for stereo

Solve $E_{structure}(I_{s1}^*, I_{s2}, D_2^*) \sim E_{Is2}(I_{s2})$ by expressing $E_{Is2}(I_{s2}) = \mathbf{f}(I_{s2}) + \mathbf{g}(I_{s2})$, where $\mathbf{f}(\cdot)$ and $\mathbf{g}(\cdot)$ contain the convex and non-convex terms respectively. Due to "windowed" operations, the above form is difficult to solve. However, with some algebraic manipulations, we can simplify it to –

$$\min_{I_{s2}} E_{Is2} = \min_{I_{s2}} \mathbf{f}(I_{s2}) + \sum_i \mathbf{g}_s(Z_i) \quad \text{s.t} \quad Z_i = \mathbf{A}_i(I_{s2}) + B_i$$

Linearly constrained optimization problem!

which can now be solved with the ADMM variant discussed in [1][2] (Linearized ADMM with "global" consensus (I_{s2}) among "local" variables (Z_i); $\mathbf{A_i}$ and B_i are some linear operators independent of I_{s2}). This gives us the solution for I_{s2} -

$$\overrightarrow{I_{s2}}^{k+1} := \left((2\mathbb{1} + 2\lambda_S \mathbb{L}_{I_{s2}} + \lambda_{SD}\alpha(\mathbb{1} - W_2)^T(\Lambda + \Lambda^T)(\mathbb{1} - W_2)) + \rho \sum_i A_i^T A_i \right)^{-1} \left((2\overrightarrow{I_{n2}} + \lambda_{SD}\alpha(\mathbb{1} - W_2)^T(\Lambda + \Lambda^T)W_1\overrightarrow{I_{s1}^*}) - \underbrace{\rho \sum_i A_i^T(\vec{B_i} - \vec{Z_i}^k + \vec{U_i}^k)}_{\text{consensus}} \right)$$

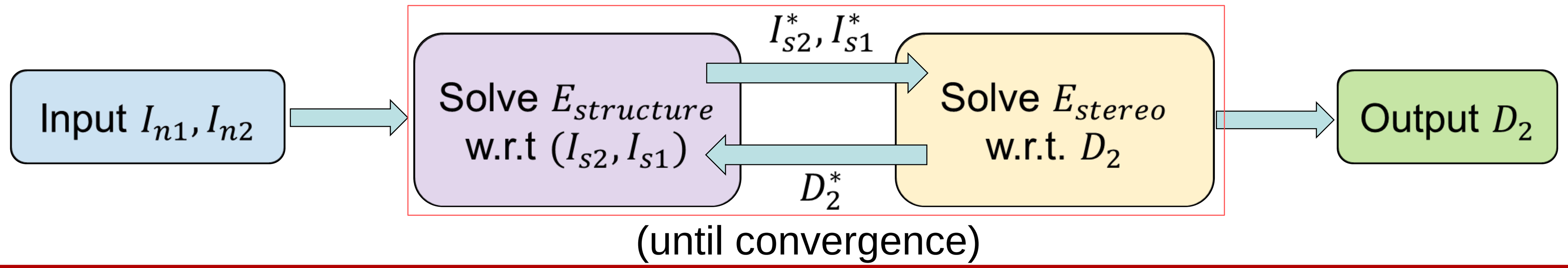
update rule for "global" I_{s2}

$$\vec{Z_i}^{k+1} := \mathrm{prox}_{\frac{1}{\rho}\mathbf{g}_s}(A_i\overrightarrow{I_{s2}}^{k+1} + \vec{B_i} + \vec{U_i}^k) \;\Big|\; \vec{U_i}^{k+1} := \vec{U_i}^k + A_i\overrightarrow{I_{s2}}^{k+1} + \vec{B_i} - \vec{Z_i}^{k+1}$$

update rule for "local" Z_i, dual U_i

Similarly, obtain the solution for $E_{structure}(I_{s1}, I_{s2}^*, D_2^*) \sim E_{Is1}(I_{s1})$

After solving $E_{Structure}(I_{s1}, I_{s2}, D_2^*)$, solve $E_{stereo}(I_{s1}^*, I_{s2}^*, D_2)$ using SGM [3], and then repeat solving the two problems until convergence. Overall process –



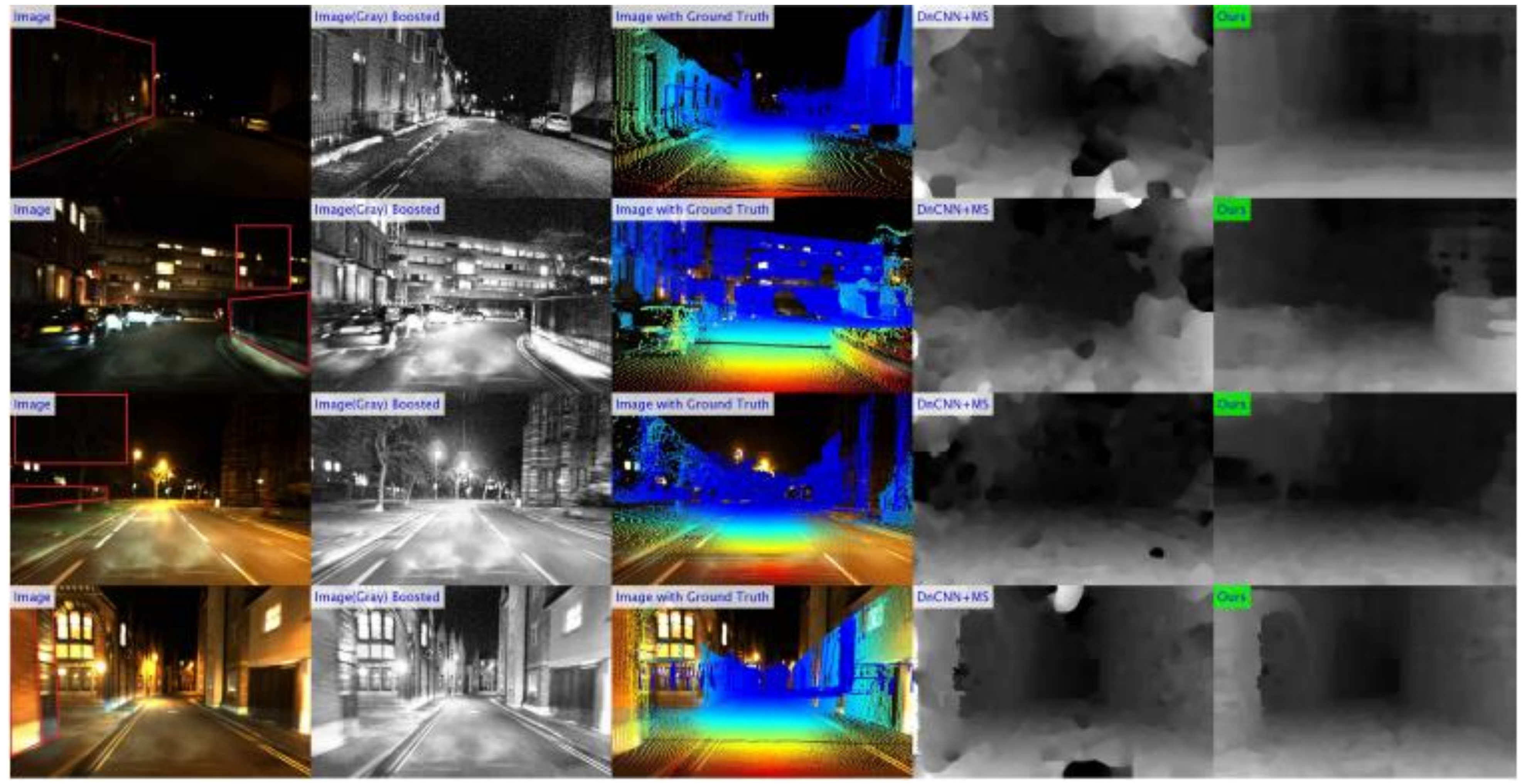
Results –

- Evaluation on Middlebury[4] (with AWGN of level $\sigma_n \in \{25, 50, 55, 60\}$)

δ	BM3D+MS				DnCNN+MS				SS-PCA				Ours			
	25	50	55	60	25	50	55	60	25	50	55	60	25	50	55	60
1px	46.45	59.55	61.29	63.35	**43.43**	55.06	57.76	58.97	51.39	60.04	66.48	69.17	47.74	**54.19**	**55.22**	**56.59**
3px	22.68	30.57	33.72	34.63	**22.04**	29.62	32.67	32.68	30.41	35.32	42.02	43.67	25.12	**29.00**	**29.45**	**30.48**
5px	**16.22**	22.01	24.17	25.07	16.82	21.53	24.36	23.94	23.14	26.07	31.48	32.93	18.21	**20.94**	**20.60**	**21.81**

- Evaluation on the Oxford RobotCar night-time dataset [5]

	DnCNN+MS					Ours				
	$\delta = 1$px	$\delta = 2$px	$\delta = 3$px	$\delta = 4$px	$\delta = 5$px	$\delta = 1$px	$\delta = 2$px	$\delta = 3$px	$\delta = 4$px	$\delta = 5$px
Set1	63.86	41.66	30.96	24.40	19.66	**58.76**	**33.75**	**23.03**	**16.99**	**12.31**
Set2	58.96	28.82	16.71	10.73	7.35	**57.76**	**28.80**	**16.10**	**10.29**	**6.82**
Set2 (f.t)	58.96	28.82	16.71	10.73	7.35	**56.45**	**26.43**	**14.54**	**9.20**	**6.08**



Summary –

- A joint structure-stereo optimization objective is proposed to solve the problem of stereo vision under night-time (or low-SNR) conditions.
- Low-PSNR but stereo-optimized images can still perform better than their high-PSNR unoptimized counterparts.

References –

[1] Parikh, N., Boyd, S., et al.: Proximal algorithms. Foundations and Trends R in Optimization 1(3) (2014) 127{239

[2] Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J., et al.: Distributed optimization and statistical learning via the alternating direction method of multipliers. Foundations and Trends R in Machine Learning 3(1) (2011) 1{122

[3] Hirschmuller, H.: Accurate and efficient stereo processing by semi-global matching and mutual information. In: CVPR'2005

[4] Scharstein, D., Hirschm•uller, H., Kitajima, Y., Krathwohl, G., Nesic, N., Wang,X., Westling, P.: High-resolution stereo datasets with subpixel-accurate ground truth. In: German Conference on Pattern Recognition, Springer (2014) 31{42

[5] Maddern, W., Pascoe, G., Linegar, C., Newman, P.: 1 Year, 1000km: The Oxford RobotCar Dataset. IJRR 36(1) (2017) 3{15

[6] Xu, L., Yan, Q., Xia, Y., Jia, J.: Structure extraction from texture via relative total variation. ACM Transactions on Graphics (TOG) 31(6) (2012) 139