Supplementary Material: Cascaded Scene Flow Prediction using Semantic Segmentation

Zhile Ren
Brown University
ren@cs.brown.edu

Deqing Sun
NVIDIA
deqings@nvidia.com

Jan Kautz
NVIDIA
jkautz@nvidia.com

Erik B. Sudderth
UC Irvine
sudderth@uci.edu

1. Learned Parameters

We split ground truth data for KITTI dataset [3] as 150 for training and 50 for validation, and learn parameters in our model at each stages of the cascade. Here, we report the learned parameters.

1.1. Refinement for Semantic Segmentation

We show in Table 1 the learned parameters for segmentation refinement model using Structural SVM training [2].

Iter1	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7
	-1.90	3.68	1.62	-3.29	0.00	0.64	0.17
	λ_8	λ_9	λ_{10}	λ_{11}	σ_{img}	σ_{disp}	
	-	-	-	-	100	3000	
Iter2	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7
	-0.45	0.92	0.36	-0.69	0.04	0.12	-0.03
	λ_8	λ_9	λ_{10}	λ_{11}	σ_{img}	σ_{disp}	
	-0.03	0.08	0.02	0.01	100	3000	
Iter3	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7
	-0.10	0.25	0.09	-0.11	0.03	0.03	-0.01
	λ_8	λ_9	λ_{10}	λ_{11}	σ_{img}	σ_{disp}	
	0.09	-0.00	0.02	100	5000	5000	

Table 1. Parameters ($\times 10^{-3}$) for segmentation refinement.

1.2. Estimating Scene Geometry, 3D Motion, 2D Optical Flow, and Flow Fusion

We learn the rest of the parameters for scene geometry modeling, 3D motion and 2D optical flow estimation, and flow fusion using grid search. The learned parameters will be shared across all stages of the cascade and is shown in Table 2.

Scene Geometry Estimation	τ_1	τ_2
	3	0.5
3D Motion Estimation	ν	
	10	
Optical Flow Estimation	η_1	η_2
	1	0.12
Flow Fusion	ω_1	ω_2
	1	0.1

Table 2. Parameters for estimating scene geometry, 3D Motion, 2D optical flow, and flow fusion.

2. Visualizing Silhouette Cost

When using cascaded prediction, we discussed in Equation (10) of section 4 on how to **recover from a poor optical flow initialization**. Specifically we adapted the idea from human pose estimation [1] and optimize the following silhouette cost function

$$\frac{1}{|B|}\sum_{i\in B}\alpha S(M)_i\cdot C(S_i^{(2)})+(1-\alpha)C(S(M)_i)\cdot S_i^{(2)}. \quad (10)$$

To provide an intuitive example, we visualized an example in Figure 1. Since the initial flow is totally wrong, the estimated motion will inevitably be problematic as well, and indeed the silhouette cost is large. For flow estimations with large silhouette errors, we replace the first term in Eq. (5) with this term. As a result, the estimated motion is reliable, leading to more accurate optical flow predictions.

3. More Qualitative Results

We demonstrate more qualitative results of our algorithm on KITTI training set [3] in Figure 2, 3 and 4.

References

[1] A. O. Balan, L. Sigal, M. J. Black, J. E. Davis, and H. W. Haussecker. Detailed human shape and pose from images. In *CVPR*, pages 1–8. IEEE, 2007. 1

[2] T. Finley and T. Joachims. Training structural svms when exact inference is intractable. In *ICML*. ACM, 2008. 1

[3] M. Menze and A. Geiger. Object scene flow for autonomous vehicles. In *CVPR*, pages 3061–3070, 2015. 1

Figure 1. Visualizing how we recover correct 3D motion from poor flow initializations by minimizing the silhouette cost.

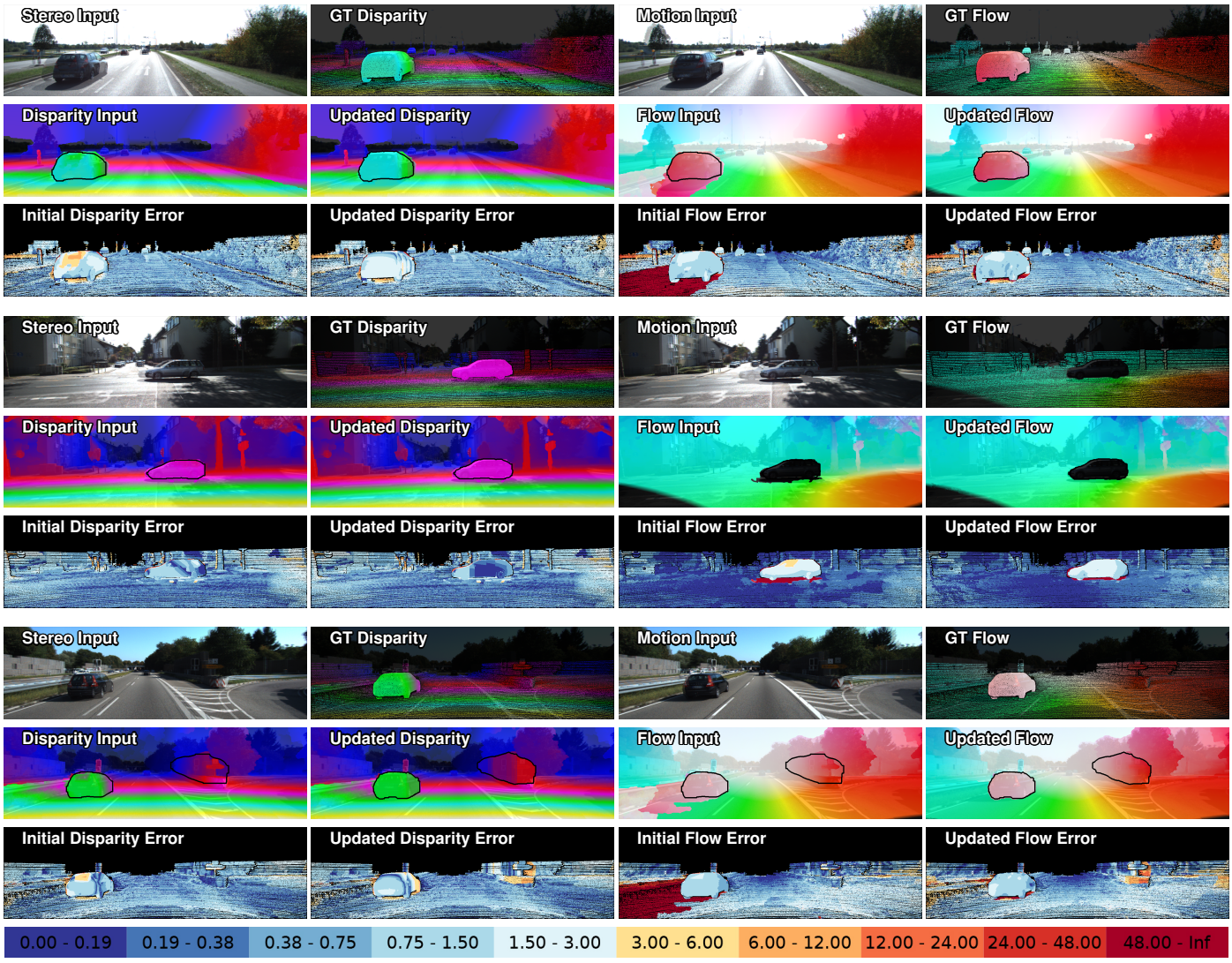

Figure 2. Visualizing the qualitative results of our algorithm in KITTI training set.

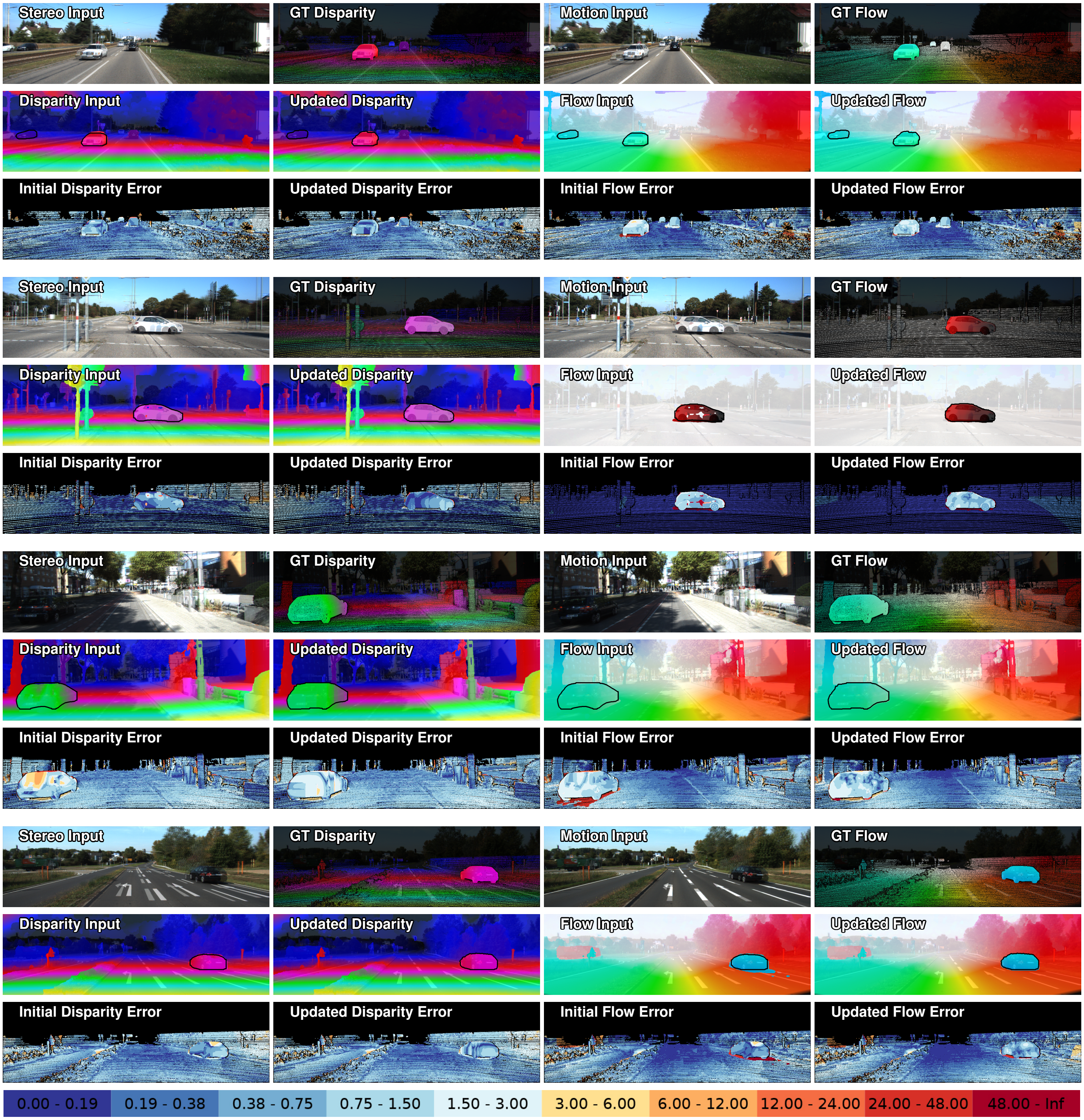

Figure 3. Visualizing the qualitative results of our algorithm in KITTI training set. (contd.)

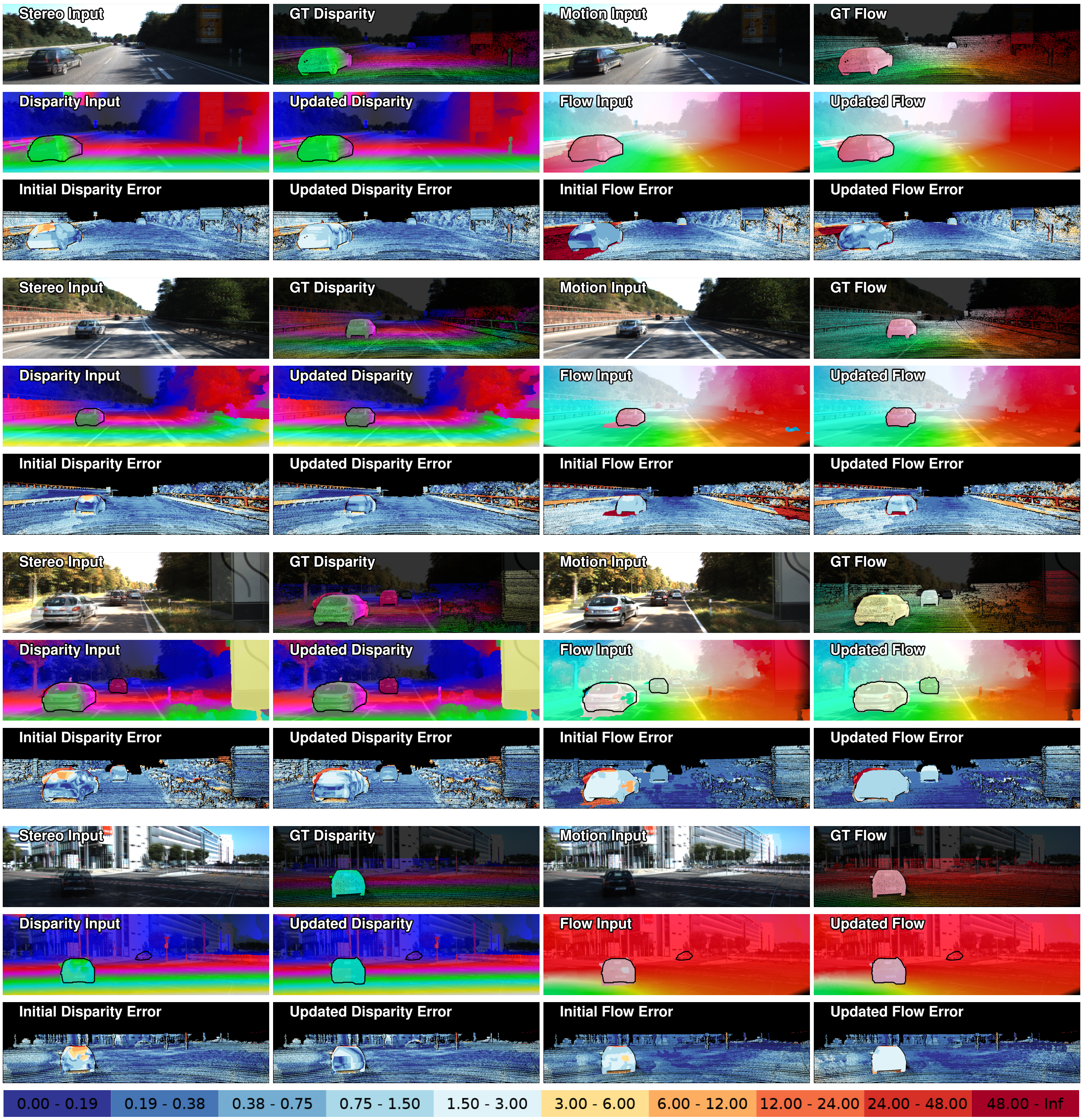

Figure 4. Visualizing the qualitative results of our algorithm in KITTI training set. (contd.)