Supplementary material for paper #529 Learning confidence measures in the wild

No Author Given

No Institute Given

In this document we provide additional experimental results concerned with paper #529 submitted to BMVC 2017 summarized in the original submission due to page limitations. To improve readability we adopt a single column format.

1 Quantitative evaluation

Table 1 extends the evaluation already reported in the paper, by showing the average AUCs of all the confidence measures involved in our self-supervised sample selection. We can observe how all of them perform worse than APKR, with the exception of WMN, achieving slightly lower AUCs in 2 out of 3 cases with the SGM algorithm. For this reason and due to the lack of space, in the paper we have included only results for APKR.

KITTI 12	CENSUS (ϵ=38.6%)			MC-CNN (ϵ=16.9%)			SGM (ϵ=9.1%)		
measure	GT	[1]	Prop.	GT	[1]	Prop.	GT	[1]	Prop.
O1 [2]	0.116	0.165	0.163	0.025	0.046	0.042	0.016	0.031	0.022
CCNN [3]	0.118	0.250	0.128	0.028	0.089	0.029	0.032	0.084	0.023
PBCP [4]	0.125	0.201	0.138	0.029	0.044	0.040	0.029	0.037	0.035
APKR [5]		0.166			0.048			0.030	
DLB [6]		0.359			0.14.5			0.078	
MED [7]		0.281			0.090			0.054	
LRC [8]		0.265			0.100			0.055	
UC [9]		0.277			0.106			0.065	
WMN [8]		0.205			0.065			0.029	
opt.		0.094			0.017			0.005	
KITTI 15	CENSUS (ϵ=35.4%)			MC-CNN (ϵ=15.4%)			SGM (ϵ=13.7%)		
measure	GT	[1]	Prop.	GT	[1]	Prop.	GT	[1]	Prop.
O1 [2]	0.109	0.172	0.147	0.031	0.059	0.046	0.021	0.038	0.027
CCNN [3]	0.113	0.266	0.120	0.036	0.102	0.035	0.044	0.072	0.029
PBCP [4]	0.122	0.209	0.151	0.035	0.053	0.047	0.031	0.035	0.037
APKR [5]		0.147			0.049			0.036	
DLB [6]		0.345			0.144			0.083	
MED [7]		0.266			0.090			0.057	
LRC [8]		0.253			0.099			0.059	
UC [9]		0.257			0.103			0.068	
WMN [8]		0.194			0.062			0.036	
opt.		0.083			0.019			0.007	
MIDD 14	CENSUS(ϵ=37.8%)			MC-CNN (ϵ=26.7%)			SGM (ϵ=26.9%)		
measure	GT	[1]	Prop.	GT	[1]	Prop.	GT	[1]	Prop.
O1 [2]	0.126	0.180	0.154	0.073	0.125	0.097	0.085	0.133	0.102
CCNN [3]	0.128	0.254	0.123	0.072	0.179	0.069	0.122	0.216	0.088
PBCP [4]	0.119	0.169	0.123	0.067	0.084	0.078	0.145	0.148	0.148
APKR [5]		0.137			0.074			0.100	
DLB [6]		0.333			0.226			0.225	
MED [7]		0.248			0.159			0.181	
LRC [8]		0.239			0.153			0.175	
UC [9]		0.254			0.166			0.189	
WMN [8]		0.158			0.076			0.099	
opt.		0.090			0.046			0.045	

Table 1. Average AUCs on the three datasets (from top to bottom: KITTI 12, KITTI 15 and MIDD 14). Evaluation of the three confidence measures with three algorithms (CENSUS, MC-CNN, SGM), trained on ground-truth data (GT), on labels obtained by SELF [1] and by the proposed method. We also include in the table a single AUC concerned with the confidence measures used by our method to select positive/negative samples, not affected at all by training labels. For each stereo algorithm we also report the average error ϵ on each dataset computed with error bound set to 3, for KITTI datasets, and set to 1 for MIDD 14.

2 Training data analysis: accuracy vs density

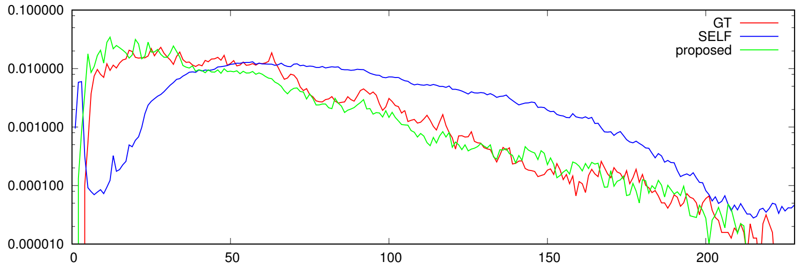
Table 2 reports detailed accuracy, density and intersection with ground-truth, for each of the eight stereo pairs. While SELF always provide a higher number of samples per image and, thus, a higher density, our method always outperforms SELF in terms of accuracy with the three confidence measures, resulting in a lower number of wrong labels predicted. On image 000180, SELF achieves a lower density compared to the other pairs. As stated by the authors [1], that sequence provides no useful cues for finding correct matches, thus the reported statistics only concerns with wrong labels.

000043	CENSUS		MC-CNN		SGM	
Method	A	D / D∩GT	A	D / D∩GT	A	D / D∩GT
SELF [1]	96.0%	**48.4% / 55.6%**	94.8%	**45.4% / 48.3%**	**91.6%**	**42.6% / 46.6%**
Prop.	**98.3%**	6.9% / 6.34%	**98.1%**	11.0% / 11.6%	86.1%	12.4% / 13.0%
000071	CENSUS		MC-CNN		SGM	
Method	A	D / D∩GT	A	D / D∩GT	A	D / D∩GT
SELF [1]	95.9%	**40.9% / 37.9%**	93.2%	**37.8% / 31.1%**	78.3%	**29.3% / 19.9%**
Prop.	**98.6%**	7.0% / 5.57%	**97.7%**	10.3% / 11.7%	**88.2%**	11.7% / 14.8%
000082	CENSUS		MC-CNN		SGM	
Method	A	D / D∩GT	A	D / D∩GT	A	D / D∩GT
SELF [1]	58.1%	**38.2% / 38.4%**	43.1%	**28.7% / 28.2%**	57.5%	**17.4% / 18.7%**
Prop.	**98.5%**	9.8% / 7.84%	**95.9%**	13.8% / 13.6%	**80.8%**	13.1% / 11.7%
000087	CENSUS		MC-CNN		SGM	
Method	A	D / D∩GT	A	D / D∩GT	A	D / D∩GT
SELF [1]	95.0%	**38.7% / 37.7%**	94.2%	**34.9% / 33.9%**	80.6%	**22.9% / 24.9%**
Prop.	**98.1%**	9.8% / 7.08%	**94.6%**	12.7% / 11.5%	**81.0%**	12.3% / 12.0%
000094	CENSUS		MC-CNN		SGM	
Method	A	D / D∩GT	A	D / D∩GT	A	D / D∩GT
SELF [1]	94.2%	**42.5% / 39.9%**	90.8%	**40.5% / 32.4%**	84.2%	**29.4% / 25.6%**
Prop.	**98.2%**	5.1% / 6.82%	**98.4%**	9.1% / 15.8%	**95.8%**	10.6% / 21.8%
000120	CENSUS		MC-CNN		SGM	
Method	A	D / D∩GT	A	D / D∩GT	A	D / D∩GT
SELF [1]	94.8%	**35.5% / 31.7%**	90.9%	**27.7% / 21.9%**	81.4%	**13.9% / 11.6%**
Prop.	**99.2%**	10.0% / 10.8%	**97.4%**	14.9% / 13.4%	**92.5%**	12.5% / 13.8%
000122	CENSUS		MC-CNN		SGM	
Method	A	D / D∩GT	A	D / D∩GT	A	D / D∩GT
SELF [1]	96.3%	**39.8% / 42.2%**	94.8%	**39.4% / 38.4%**	91.9%	**39.5% / 38.3%**
Prop.	**99.3%**	7.2% / 12.0%	**99.2%**	11.3% / 19.4%	**95.6%**	13.6% / 22.7%
000180	CENSUS		MC-CNN		SGM	
Method	A	D / D∩GT	A	D / D∩GT	A	D / D∩GT
SELF [1]	93.5%	**26.1% / 19.7%**	84.4%	**17.1% / 11.5%**	56.3%	6.60% / 4.70%
Prop.	**96.7%**	10.7% / 7.5%	**95.1%**	14.1% / 11.8%	**88.5%**	**12.8% / 13.3%**

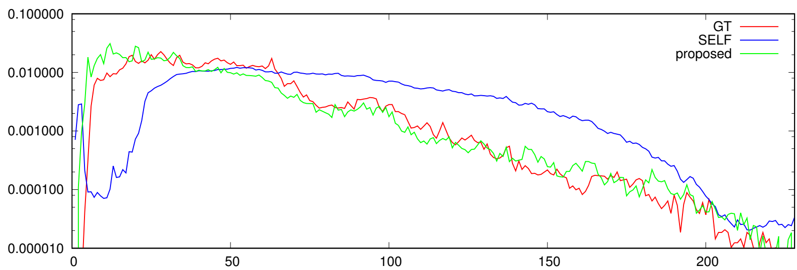
Table 2. Analysis of training labels inferred on eight sequences of KITTI 12. For SELF [1] and the proposed method we report the accuracy A for the predicted labels (computed for points with available ground-truth), the average density D on the eight sequences, the intersection between the density of labels inferred by the two methods and the eight images with ground-truth (D∩GT). The average density of KITTI 12 ground-truth data on the eight images is 19.5%.

3 Training data analysis: distribution

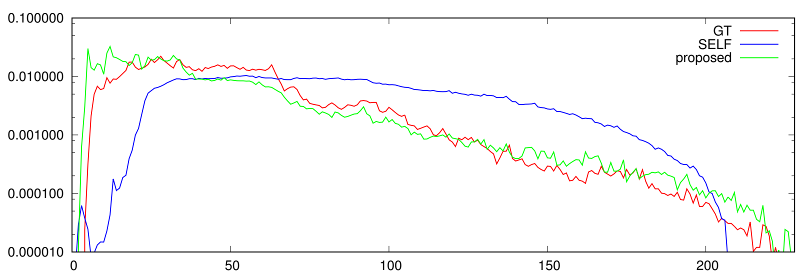
Figure 1 and 2 report, respectively, the distribution of *positive* and *negative* samples for the entire disparity range concerned with positive and negative samples extracted by SELF [1] and by the proposed method with disparity maps provided by three stereo algorithms CENSUS, MC-CNN [10] and SGM [11]. The figure also reports the distribution of positive and negative samples for ground-truth data (independent of the stereo stereo algorithm deployed). While SELF and the proposed method generate a substantially similar distribution of negative samples, the distribution of positive samples is very different. In particular, we can observe how the proposed method generates a distribution very similar to the ground-truth, while SELF provides very few positive samples at lower disparity values (i.e., for farther 3D points). Being the three state-of-the-art confidence measures O1 [2], CCNN [3] and PBCP [4] based on disparity map analysis, this unbalanced availability of positive and negative samples at low disparities greatly affects the training procedure, leading the confidence measures trained with SELF labels to assign low scores to farther points. This is particularly evident by looking at confidence maps outcome of different methods, reported qualitatively in the next section.



(a)

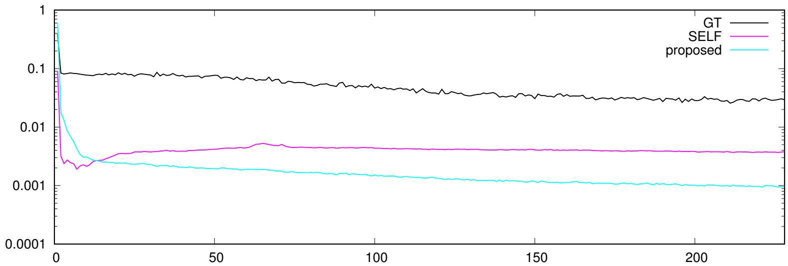


(b)

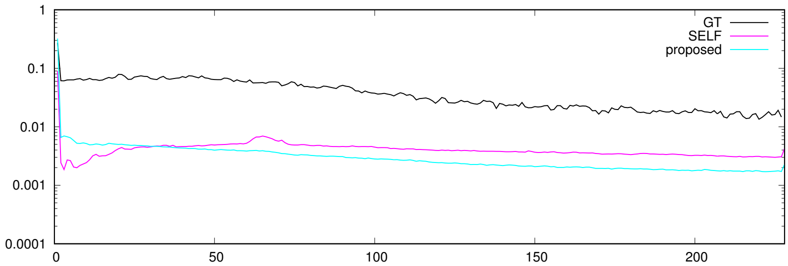


(c)

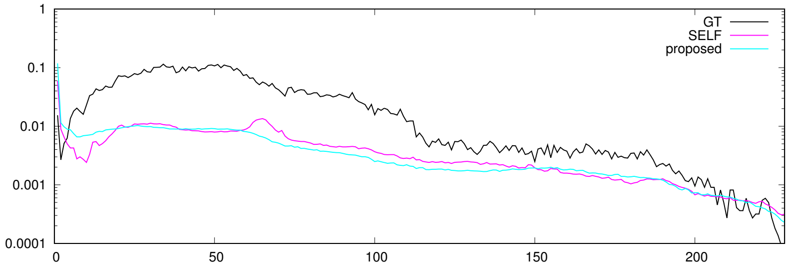
Fig. 1. Distribution of *positive* samples on the disparity range, according to ground-truth data (red), SELF [1] (blue) and proposed (green) methods. Results on CENSUS (a), MC-CNN [10] (b) and SGM [11] (c) data.



(a)



(b)

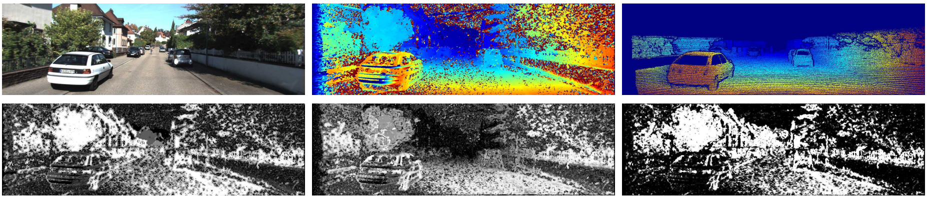


(c)

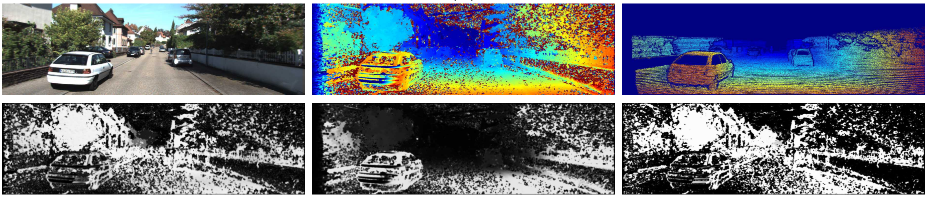
Fig. 2. Distribution of *negative* samples on the disparity range, according to ground-truth data (black), SELF [1] (purple) and proposed (cyan) methods. Results on CENSUS (a), MC-CNN [10] (b) and SGM [11] (c) data.

4 Qualitative results

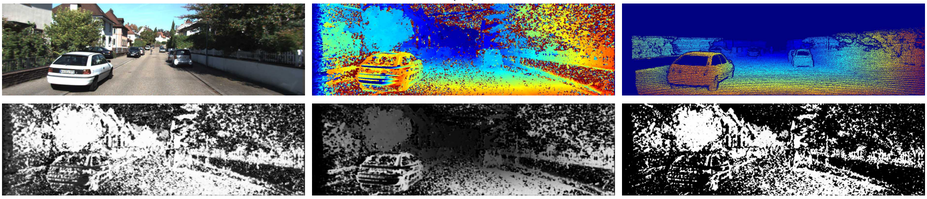
To conclude, we report qualitative results by showing some examples of confidence maps obtained by training the three confidence measures with the proposed method, with SELF [1] and with ground-truth data. Figures 3, 4 and 5 depict maps from stereo pair 000006 of the KITTI 2012 dataset [12], respectively for CENSUS, MC-CNN and SGM algorithms. Figures 6, 7 and 8 depict maps from stereo pair *Adirondack* of the Middlebury 2014 dataset [13], respectively for CENSUS, MC-CNN and SGM algorithms. We can observe how measures trained with SELF often assign low confidence to farther points on the scene, as result of the unbalanced distribution of samples highlighted in the previous section.



(a)

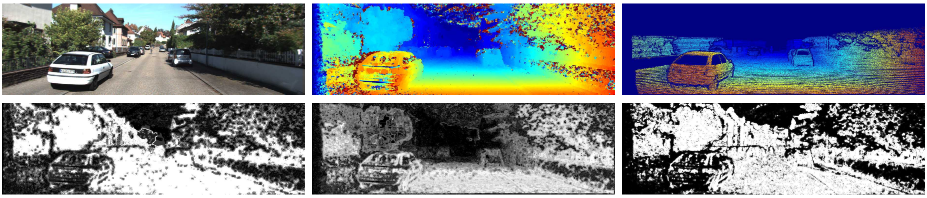


(b)

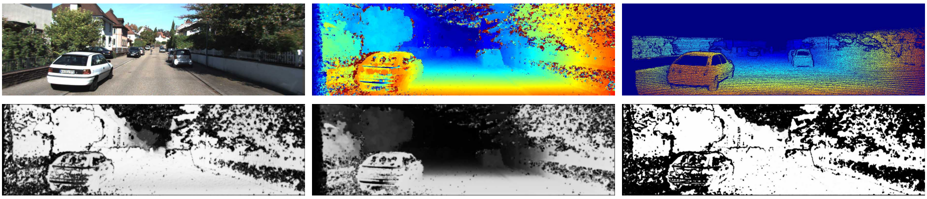


(c)

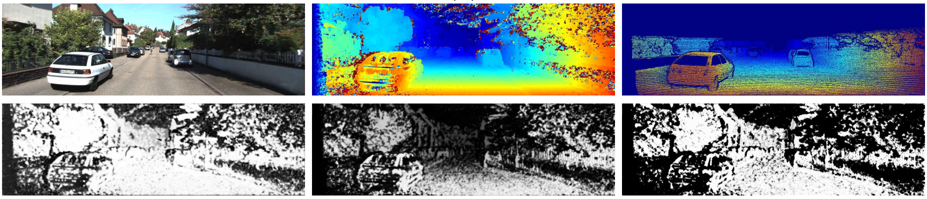
Fig. 3. Qualitative results on KITTI 000006 stereo pair, CENSUS algorithm. (a) O1 [2], (b) CCNN [3], (c) PBCP [4]. For each measure, we report reference image, disparity map, ground-truth data and confidence maps obtained, respectively, by training on ground-truth data, using SELF [1] and the proposed method.



(a)

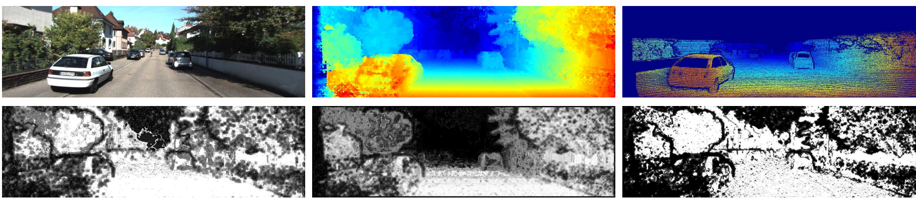


(b)

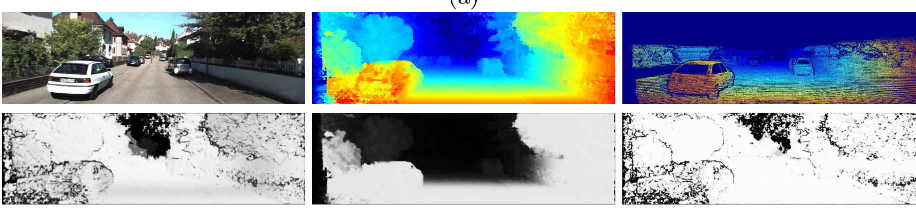


(c)

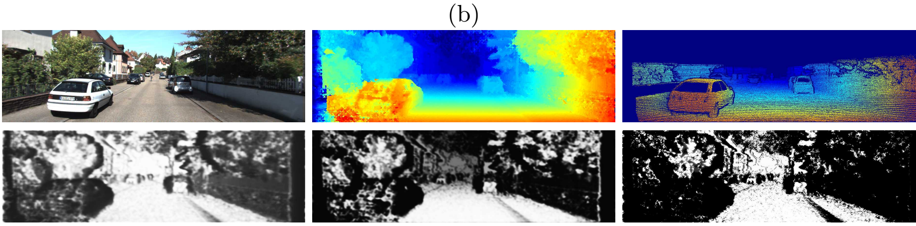
Fig. 4. Qualitative results on KITTI 000006 stereo pair, MC-CNN algorithm. (a) O1 [2], (b) CCNN [3], (c) PBCP [4]. For each measure, we report reference image, disparity map on top, confidence maps obtained, respectively, by training on ground-truth data, using SELF [1] and the proposed method on bottom.



(a)



(b)



(c)

Fig. 5. Qualitative results on KITTI 000006 stereo pair, SGM algorithm. (a) O1 [2], (b) CCNN [3], (c) PBCP [4]. For each measure, we report reference image, disparity map on top, confidence maps obtained, respectively, by training on ground-truth data, using SELF [1] and the proposed method on bottom.

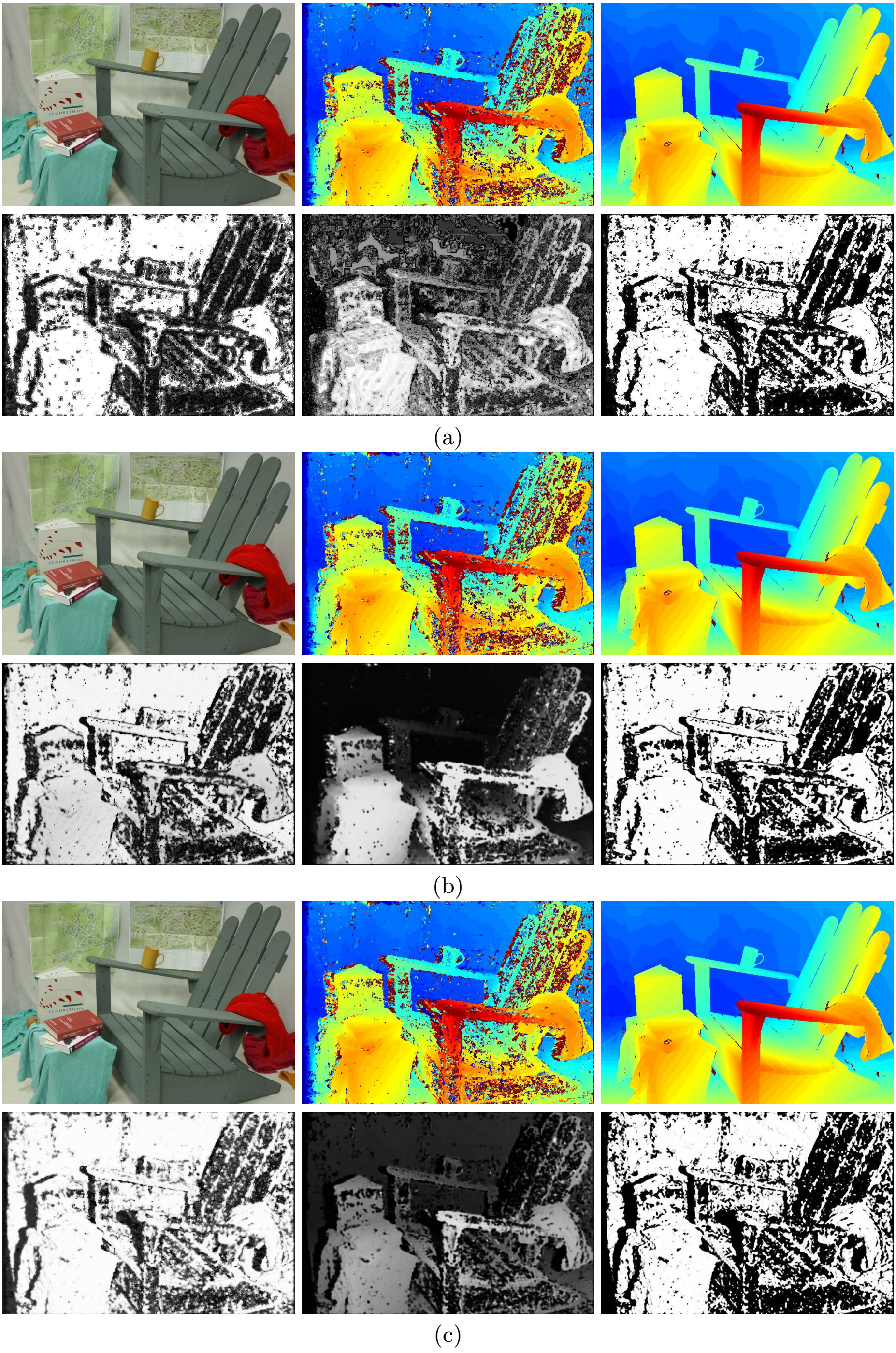


Fig. 6. Qualitative results on Middlebury *Adirondack* stereo pair, CENSUS algorithm. (a) O1 [2], (b) CCNN [3], (c) PBCP [4]. For each measure, we report reference image, disparity map, ground-truth data on top, confidence maps obtained, respectively, by training on ground-truth data, using SELF [1] and the proposed method on bottom.

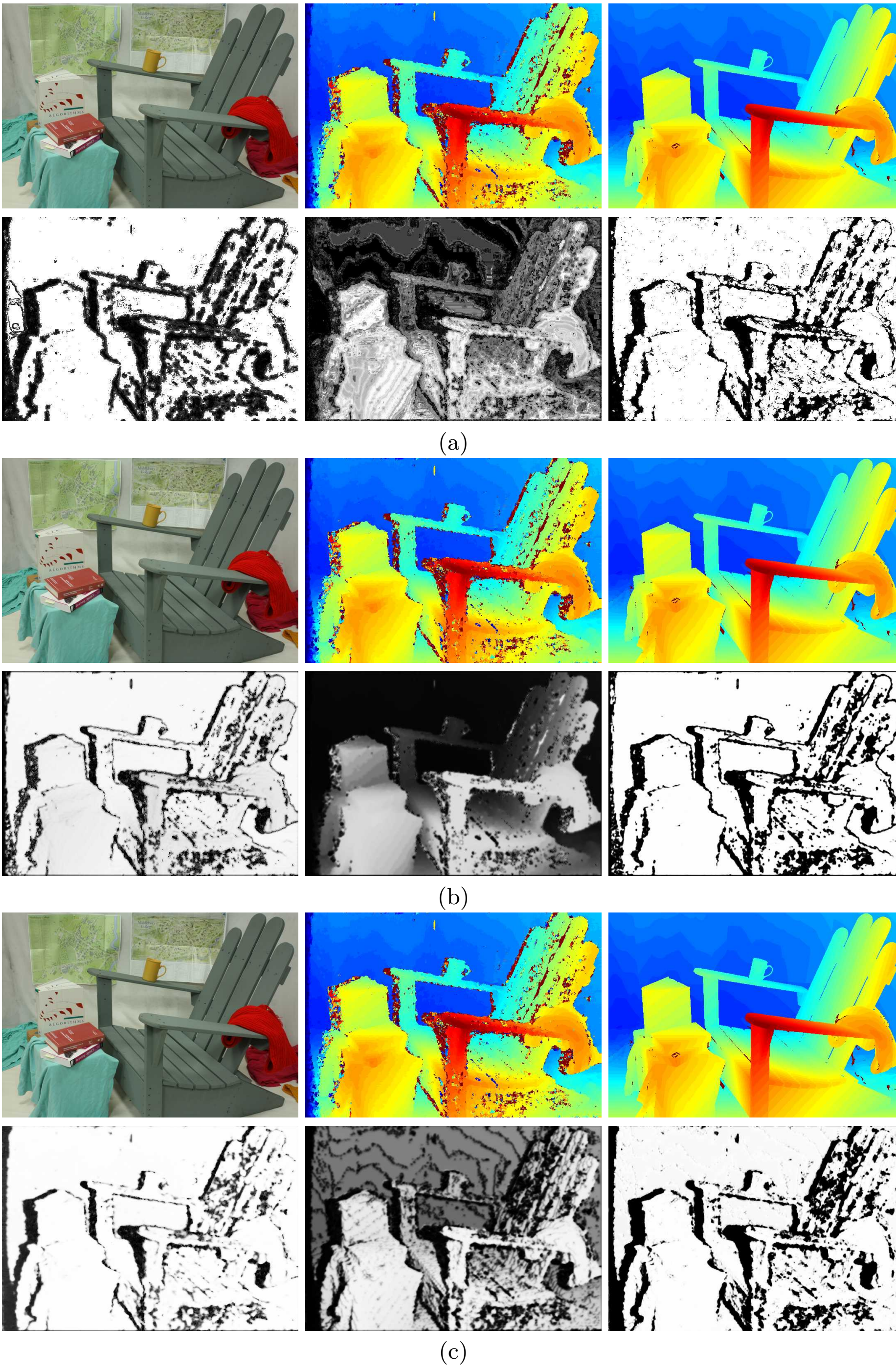
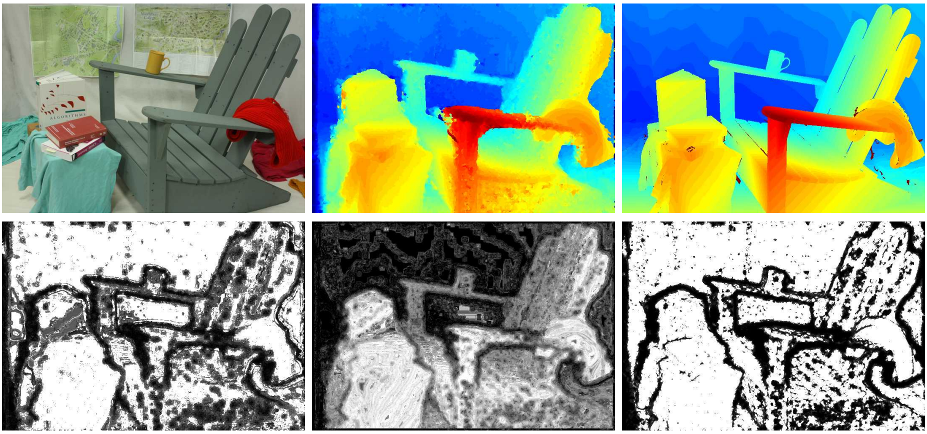
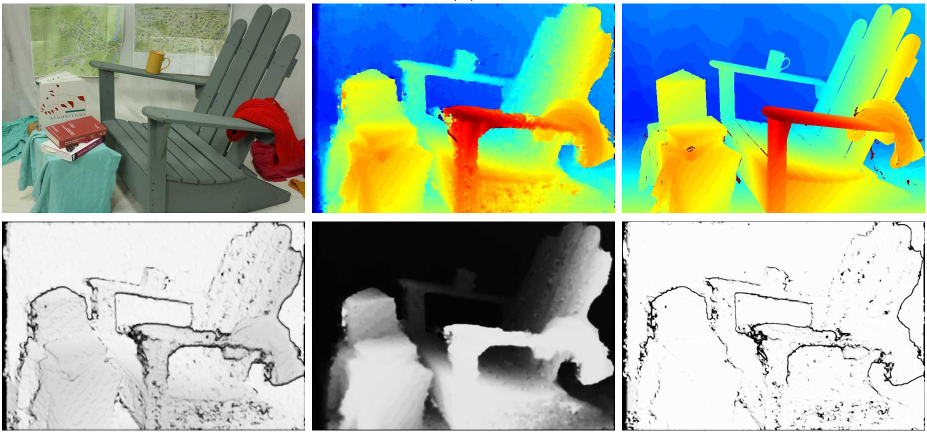


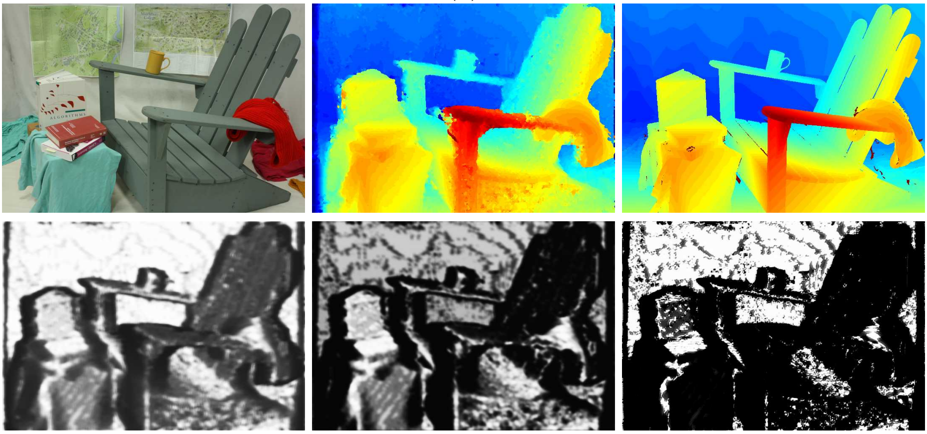
Fig. 7. Qualitative results on Middlebury *Adirondack* stereo pair, MC-CNN algorithm. (a) O1 [2], (b) CCNN [3], (c) PBCP [4]. For each measure, we report reference image, disparity map, ground-truth data on top, confidence maps obtained, respectively, by training on ground-truth data, using SELF [1] and the proposed method on bottom.



(a)



(b)



(c)

Fig. 8. Qualitative results on Middlebury *Adirondack* stereo pair, SGM algorithm. (a) O1 [2], (b) CCNN [3], (c) PBCP [4]. For each measure, we report reference image, disparity map, ground-truth data and confidence maps obtained, respectively, by training on ground-truth data, using SELF [1] and the proposed method.

References

1. Mostegel, C., Rumpler, M., Fraundorfer, F., Bischof, H.: Using self-contradiction to learn confidence measures in stereo vision. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2016)
2. Poggi, M., Mattoccia, S.: Learning a general-purpose confidence measure based on o(1) features and asmarter aggregation strategy for semi global matching. In: Proceedings of the 4th International Conference on 3D Vision, 3DV. (2016)
3. Poggi, M., Mattoccia, S.: Learning from scratch a confidence measure. In: Proceedings of the 27th British Conference on Machine Vision, BMVC. (2016)
4. Seki, A., Pollefeys, M.: Patch based confidence prediction for dense disparity map. In: British Machine Vision Conference (BMVC). (2016)
5. Kim, S., g. Yoo, D., Kim, Y.H.: Stereo confidence metrics using the costs of surrounding pixels. In: 2014 19th International Conference on Digital Signal Processing. (Aug 2014) 98–103
6. Park, M.G., Yoon, K.J.: Leveraging stereo matching with learning-based confidence measures. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (June 2015)
7. Spyropoulos, A., Komodakis, N., Mordohai, P.: Learning to detect ground control points for improving the accuracy of stereo matching. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE (2014) 1621–1628
8. Hu, X., Mordohai, P.: A quantitative evaluation of confidence measures for stereo vision. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI) (2012) 2121–2133
9. Di Stefano, L., Marchionni, M., Mattoccia, S.: A fast area-based stereo matching algorithm. Image and vision computing **22**(12) (2004) 983–1005
10. Zbontar, J., LeCun, Y.: Stereo matching by training a convolutional neural network to compare image patches. Journal of Machine Learning Research **17** (2016) 1–32
11. Hirschmuller, H.: Stereo processing by semiglobal matching and mutual information. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI) **30**(2) (feb 2008) 328–341
12. Geiger, A., Lenz, P., Stiller, C., Urtasun, R.: Vision meets robotics: The kitti dataset. Int. J. Rob. Res. **32**(11) (sep 2013) 1231–1237
13. Scharstein, D., Szeliski, R.: High-accuracy stereo depth maps using structured light. In: Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR'03, Washington, DC, USA, IEEE Computer Society (2003) 195–202