Pacific Graphics 2019
C. Theobalt, J. Lee, and G. Wetzstein
(Guest Editors)

Volume 38 (2019), Number 7

Appendix for: Generic interactive pixel-level image editing

A. NYU depth v_2 visual results

As our paper mentioned before, the size of the the NYU depth V_2 input is 460*620 and we segment the input into about 300 superpixels. The color of the strokes stand for different meaningful constraint. Firstly, the assignment constraint is input from blue to red for near to far. Secondly, the depth equal constraint is input by yellow. Thirdly, the depth gradient equal is input by green.

B. NYU Depth V_2 quantative results

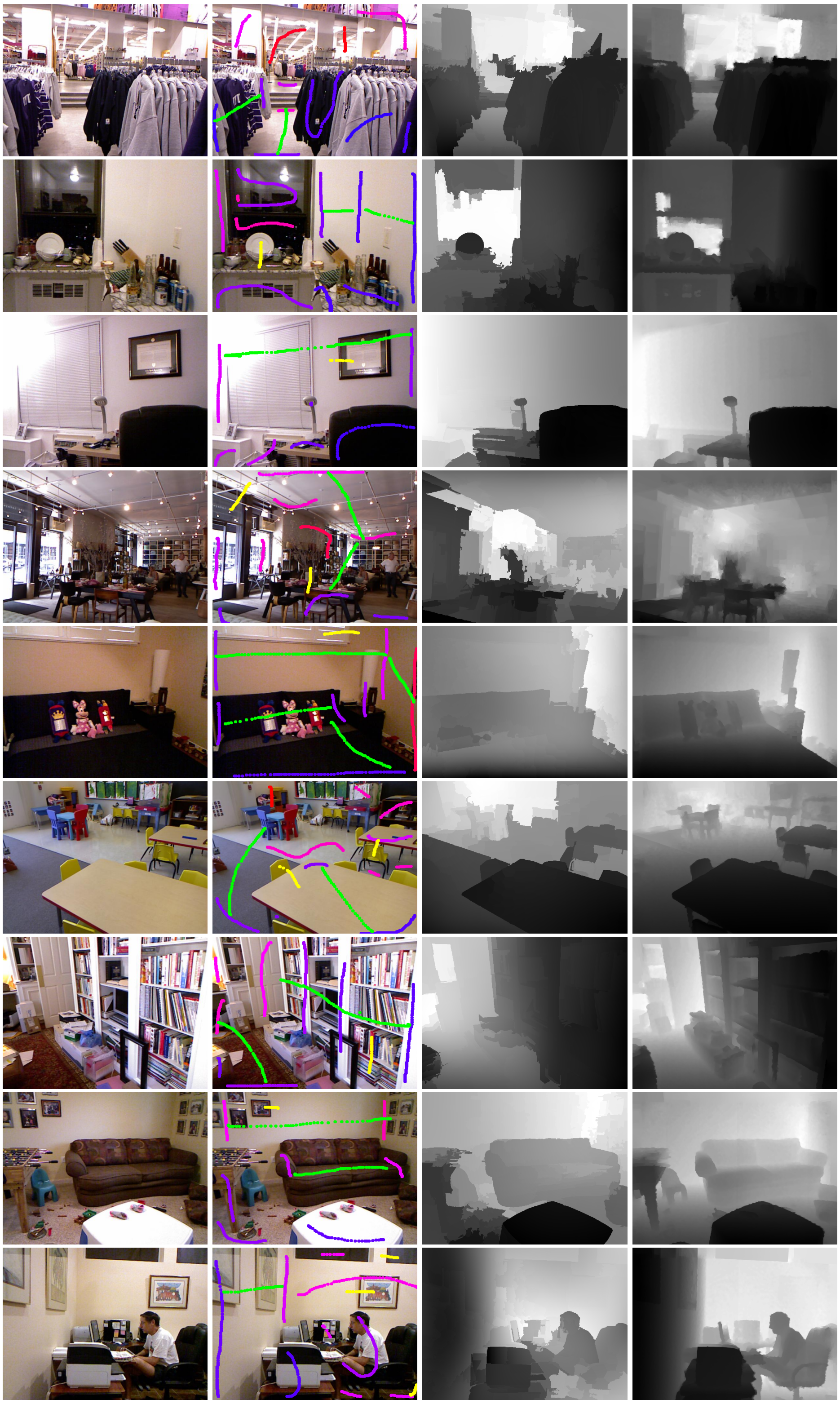
For each image, the required user strokes are demonstrated in the fifth column of Figure 4. As our seed value brushes (that set specific depth values in this case) are subjectively specified to be ("near", "far") for this experiment, we configure them to yield specific ground values (in order to avoid unfair inaccuracies due to subjective), and the rest of brushes work normally. For the generation of each result, the user has required a sparse set of user strokes. Usually, we totally just need stroke the image from 4 times to 10 times from the different brushes. Indeed, it is unfair to compare our interactive method with the automatic methods. However, the other interactive methods produce relative depth not the real depth value which cannot be evaluated on this dataset. Instead of proving to be comparable to previous work [LSH14], we use these comparison only to demonstrate that our approach has the ability to get high accuracy and reasonably good results when high quality user inputs (specific ground values) are available. Given this, the metrics are not entirely comparable but they are still an acceptable comparison available given the circumstances.

Table 1: Comparison with previous work with several error and accuracy metrics over the NYU V_2 dataset.

	Error (Lower is better)			Accuary(Higer is better)		
	rel	log10	rms	$\theta < 1.25$	$\theta < 1.25^2$	$\theta < 1.25^3$
[SSN09]	0.349	-	1.1214	0.447	0.745	0.897
[KLK12]	0.35	0.131	1.2	-	-	-
[LSH14]	0.335	0.127	1.06	-	-	-
[LSP14]	-	-	-	0.542	0.829	0.941
[EPF14]	0.214	-	0.877	0.614	0.888	0.972
[LSL14]	0.230	0.095	0.824	0.614	0.883	0.972
[EF15]	0.158	-	0.641	0.769	0.950	0.988
Ours	0.105	0.049	0.433	0.860	0.971	0.993

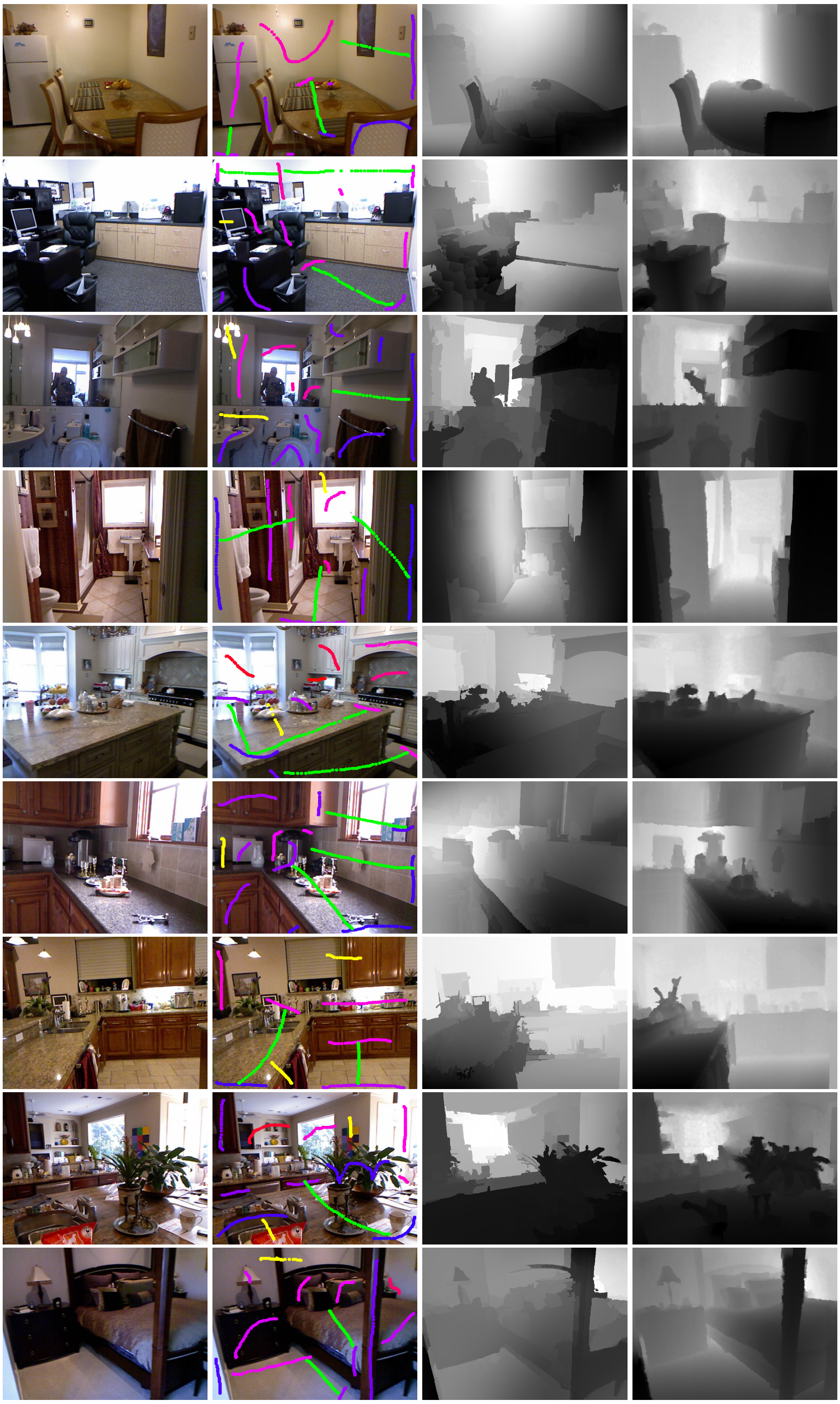
we utilize the errors and accuracy metrics presented in [LSH14] to achieve the quantitative evaluations of our method. We compare our method with many present methods as shown in Table 1. The "rel" in Table 1 describes the average relative error which is defined by $\frac{1}{N}\sum_{(x,y)}\frac{|L^*(x,y)-L(x,y)|}{L(x,y)}$. The "log10" describes the average logarithmic error which is defined by $\frac{1}{N}\sum_{(x,y)} log_{10}(L^*(x,y)) - log_{10}(L(x,y))$. The "rms" is the root mean square error which is defined by $\sqrt{\frac{1}{N}\sum_{(x,y)}(L^*(x,y)-L(x,y))^2}$. The L and L^* for these errors separately describe the estimated additional per-pixel values. In addition, the $\theta < x$ for the accuracy describes the rate of pixels (w.r.t. total pixels) whose θ metric is below x, and θ is defined by $\max(\frac{L^*(x,y)}{L(x,y)}, \frac{L(x,y)}{L^*(x,y)})$. In most cases, as we do not have access to the individual per-scene results or to the corresponding source code, we include the error metric data reported in the corresponding paper (leaving it blank when it is not available). When such data is available [LSH14] (and for our approach), we calculate the metric as the average metric for the same 30 scenes we have tested with our approach.

As shown in Table 1, by owning smaller errors and bigger accuracy, our method demonstrates its efficiency in producing the real depth values of the scenes from **NYU V_2 dataset**. However, those result metrics are not entirely comparable: as stated above, for most of the previous work the data does not include exactly the same 34 scenes. On the other hand, our results are based on a reasonably sparse user interaction. A more precise fine-tuned interaction would get more extra accuracy for our approach. Nevertheless, this experiment proves that our approach can be considerable to be on comparison with the state of the art on depth estimation and that it is able to get reasonably accurate additional per-pixel values with very simple user interaction. The rest of the evaluations and the related results are shown in the supplementary Table 2.



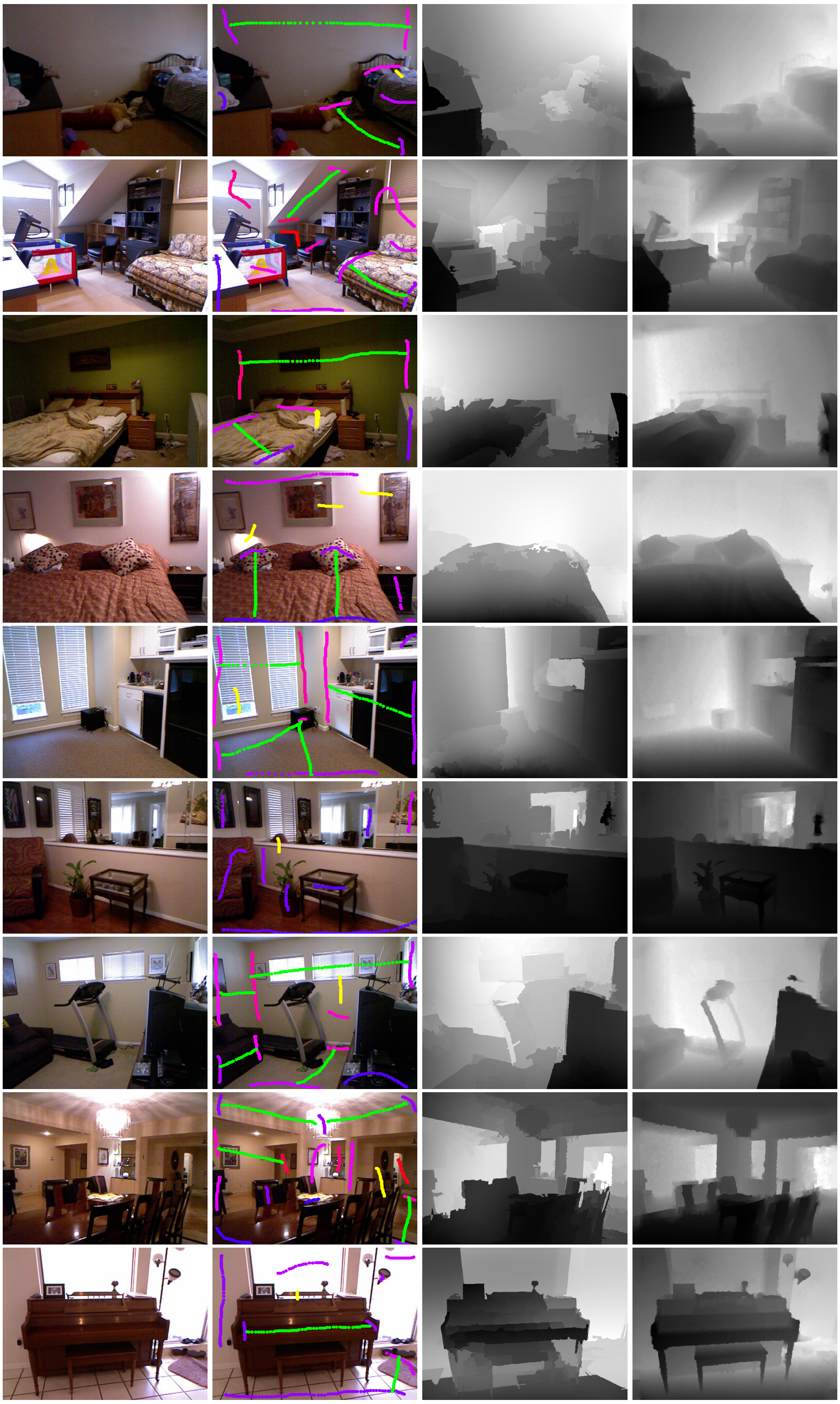
(a) Input (b) User Input (c) Our Depth (d) Ground Truth

Figure 1: NYU V_2 $1-9$



(a) Input (b) User Input (c) Our Depth (d) Ground Truth

Figure 2: NYU V_2 $10-18$



(a) Input (b) User Input (c) Our Depth (d) Ground Truth

Figure 3: NYU V_2 $19-27$

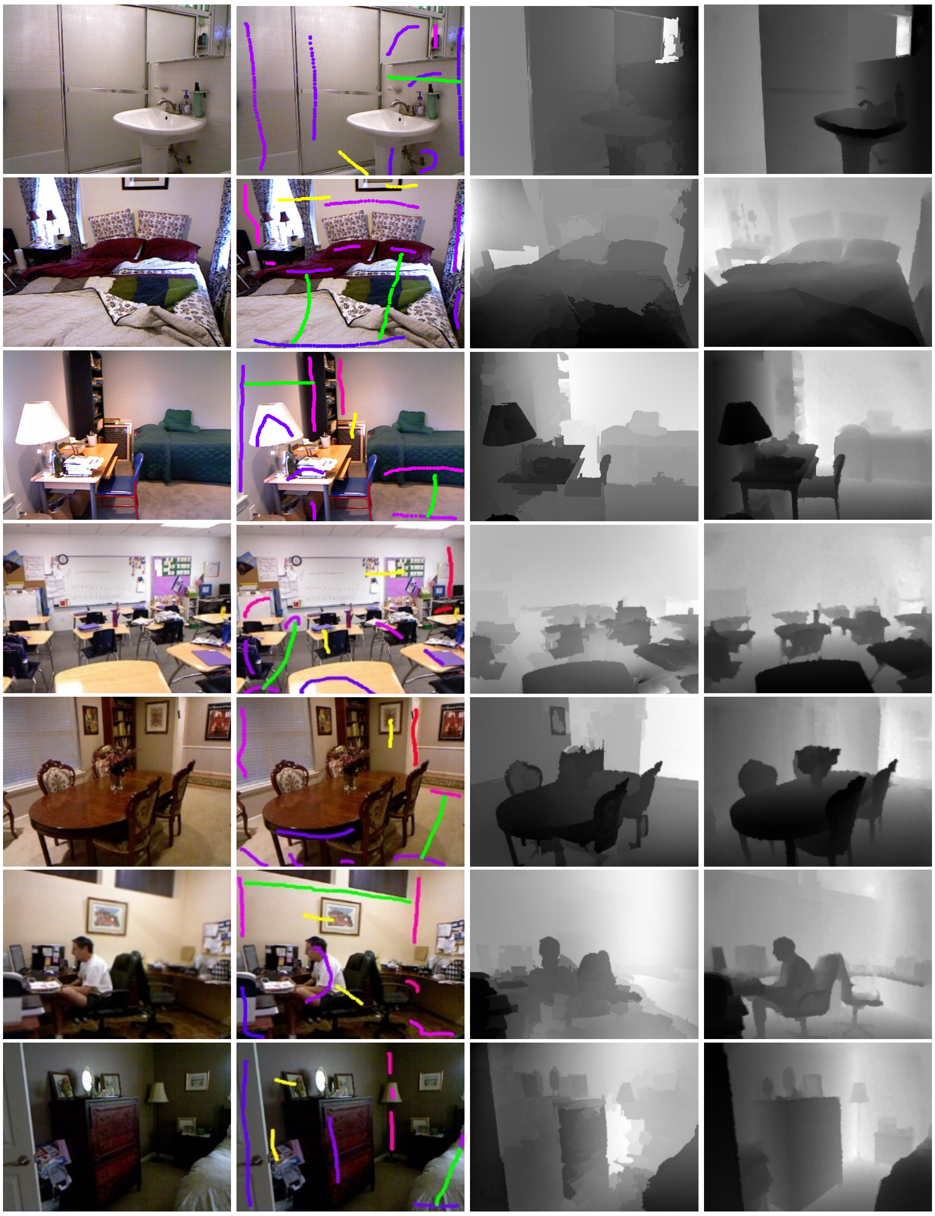


Figure 4: NYU V_2 $28-34$

Table 2: NYU Depth V_2 quantative results

Sequence	rel	log10	rms	$\theta < 1.25$	$\theta < 1.25^2$	$\theta < 1.25^3$
1	0.179	0.087	1.311	0.739	0.875	0.946
2	0.151	0.057	0.492	0.808	0.909	0.986
3	0.044	0.021	0.171	0.940	0.984	0.996
4	0.126	0.056	0.892	0.831	0.982	1.000
5	0.082	0.035	0.184	0.919	1.000	1.000
6	0.115	0.050	0.426	0.834	0.996	0.999
7	0.092	0.043	0.246	0.834	0.996	0.999
8	0.101	0.046	0.330	0.932	0.996	0.999
9	0.113	0.526	0.280	0.828	0.973	0.995
10	0.131	0.051	0.367	0.822	0.979	1.000
11	0.096	0.045	0.395	0.859	0.985	0.999
12	0.109	0.048	0.377	0.862	0.959	0.986
13	0.139	0.059	0.427	0.781	0.955	0.998
14	0.146	0.064	0.588	0.750	0.966	0.993
15	0.114	0.049	0.223	0.872	0.988	1.000
16	0.084	0.039	0.359	0.901	0.978	1.000
17	0.184	0.068	0.991	0.789	0.914	0.964
18	0.120	0.050	0.414	0.850	0.981	0.991
19	0.076	0.033	0.235	0.974	0.998	1.000
20	0.092	0.042	0.462	0.886	0.986	1.000
21	0.780	0.036	0.387	0.884	0.992	1.000
22	0.078	0.033	0.220	0.940	0.994	1.000
23	0.055	0.024	0.219	0.993	1.000	1.000
24	0.168	0.065	0.839	0.766	0.937	0.989
25	0.145	0.070	0.609	0.727	0.933	0.988
26	0.171	0.075	0.993	0.728	0.927	0.976
27	0.062	0.027	0.164	0.986	1.000	1.000
28	0.064	0.027	0.147	0.970	0.994	0.998
29	0065	0.028	0.185	0.969	0.999	1.000
30	0.060	0.026	0.226	0.971	0.998	1.000
31	0.112	0.068	0.600	0.754	0.940	0.991
32	0.088	0.045	0.388	0.807	0.935	0.999
33	0.086	0.025	0.357	0.855	0.990	1.000
34	0.054	0.035	0.231	0.840	0.996	0.999
average	0.105	0.049	0.433	0.860	0.971	0.993