Supplementary Material

Efficient High-Resolution Stereo Matching using Local Plane Sweeps
CVPR 2014

Sudipta N. Sinha
Microsoft Research
sudipsin@microsoft.com

Daniel Scharstein
Middlebury College
schar@middlebury.edu

Richard Szeliski
Microsoft Research
szeliski@microsoft.com

In the supplementary material we first provide more detail and specific parameter values for the cost functions described in the paper and then present the complete set of experimental results.

NCC-based matching costs

The matching cost based on normalized cross correlation (Section 5 of the paper) is defined as follows:

$$NCC(x,y,d) = \frac{\sum_i (u_i - \overline{u})(v_i - \overline{v})}{\sqrt{\sum_i (u_i - \overline{u})^2 \sum_i (v_i - \overline{v})^2 + \epsilon^2}},$$

where vectors $\mathbf{u}$ and $\mathbf{v}$ with $u_i, v_i \in [0, 255]$ denote pixel intensities of 3×3 patches being compared. We use $\epsilon = 10$ to downweight textureless regions with low variance. The NCC score is truncated and inverted to obtain a per-pixel matching cost $C(x,y,d) = 1 - \max(0, NCC(x,y,d))$. In our implementation, these per-pixel matching costs C and the pairwise terms V_{pq} (Equation 1) are multiplied by 255 and rounded to the nearest integers. This allows us to efficiently run SGM for the local plane sweeps using unsigned shorts without any floating point computation.

Cost map for identifying in-range disparities

In Section 5, Equation 3, we introduced a per-pixel cost map $U(p)$ that can be computed from D^*, the disparity map recovered by solving a local plane sweep problem:

$$U(p) = \lambda_R R^*(p) + \lambda_C C^*(p) + \lambda_J J^*(p).$$

The first term $R^*(p) = |I_p - I'_p|$ is the absolute intensity residual between the left image I and the right image I' warped using D^*, with $I_p, I'_p \in [0, 255]$. The second term $C^*(p)$ is the NCC-based cost weighted by the gradient magnitude at p. The third term $J^*(p)$ encodes disparity steps between adjacent pixels in D^* and is defined as

$$J^*(p) = s_J \max\left(0, D^*_x(p)^2 + D^*_y(p)^2 - 2\right),$$

where $D^*_x(p)$ and $D^*_y(p)$ are disparity changes at p for a unit horizontal and vertical step respectively. The disparity jump map $J^*(p)$ is non-zero only when the disparity steps in either direction exceed one pixel. The parameter $s_J = 10$ is used to scale J^* to be in the same range as R^* and C^*. For the relative weights, as stated in the paper, we use $\lambda_R = 0.25$, $\lambda_C = 0.25$ and $\lambda_J = 0.5$.

Pairwise term in global energy function

In Section 7, Equation 4, we defined the following energy function for the global optimization stage of our method:

$$E(L) = \sum_p U_p(l_p) + \sum_{p,q} V_{pq}(l_p, l_q).$$

The unary term $U_p(l_p)$ is defined in the paper; for pairwise term $V_{pq}(l_p, l_q)$ we use a contrast-sensitive Potts model

$$V_{pq}(l_p, l_q) = \begin{cases} 0 & \text{if } l_p = l_q \\ w(1 + \alpha e^{-|\Delta I|/\sigma_I}) & \text{otherwise} \end{cases} . \quad (1)$$

Here $w = 25$, $\sigma_I = 8$, $\alpha = 10$, and $\Delta I = I_p - I_q$ is the intensity difference for pixels p and q, with $I_p, I_q \in [0, 255]$.

Additional Results

The additional results are presented in the following order.

- Figures 1 and 2 show error histograms and scatter plots measuring average accuracy versus runtime for all five methods. The top rows are identical to Figures 4 and 5 in the paper and show results for error threshold $t = 1.0$. The bottom rows show the results for $t = 2.0$. It can be seen that the results are qualitatively similar.
- Table 1 lists the complete set of accuracies and runtimes in numerical form.
- Figure 3 shows the full set of plots illustrating the effect of different number of rounds on our method. The first of these six plots appears in Figure 6b in the paper.
- Figures 4, 5, and 6 show images, ground-truth disparities, and the disparity maps produced by the best-performing three methods for the complete sets of test images.
- Figure 7 shows example disparity and error maps for all five methods for the *Motorcycle* pair.

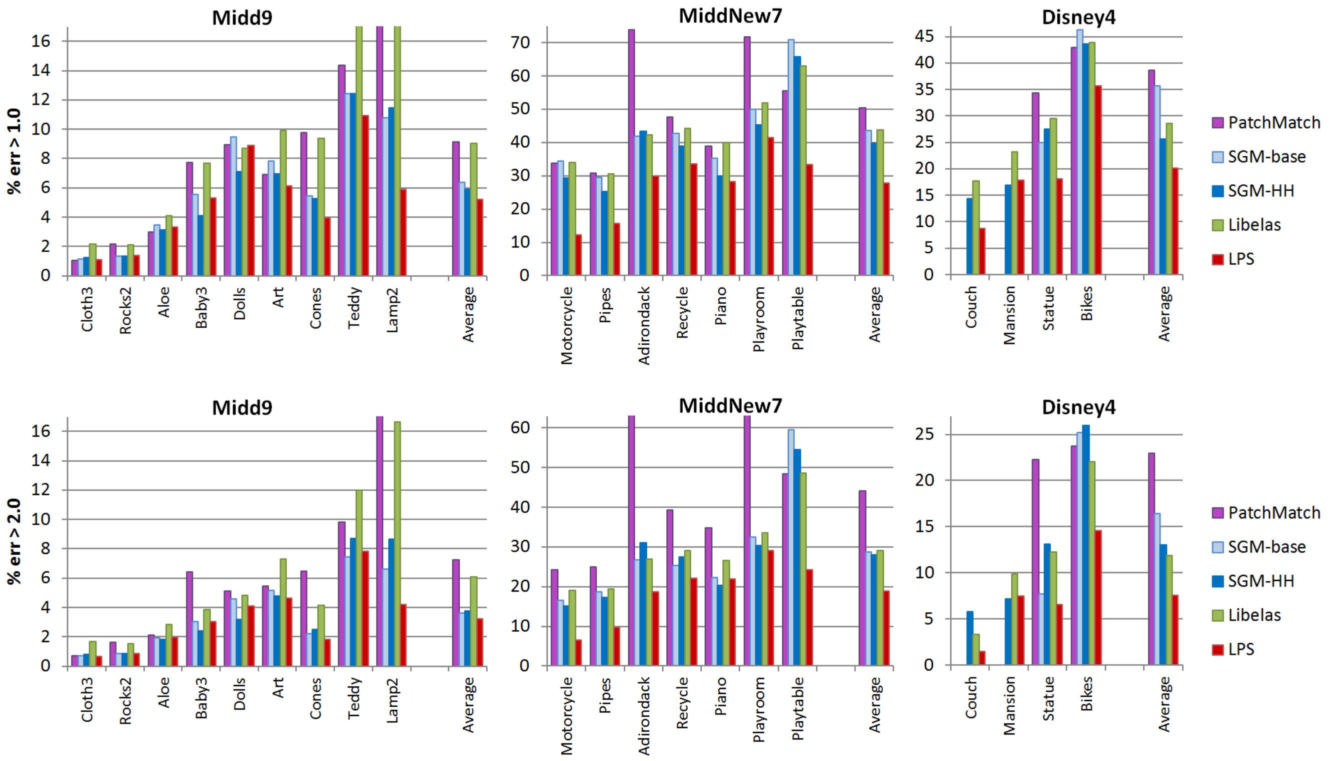

Figure 1. Error rates (% bad pixels) for error thresholds $t = 1.0$ (top) and $t = 2.0$(bottom) on the three sets of test images Midd9, MiddNew7 and Disney4, where the average image resolution is 1.7 MP, 5.5 MP and 10 MP respectively. Our method yields the lowest average errors on all three sets under both thresholds. Note that PatchMatch and SGM-base could not be run on the two largest Disney4 datasets.

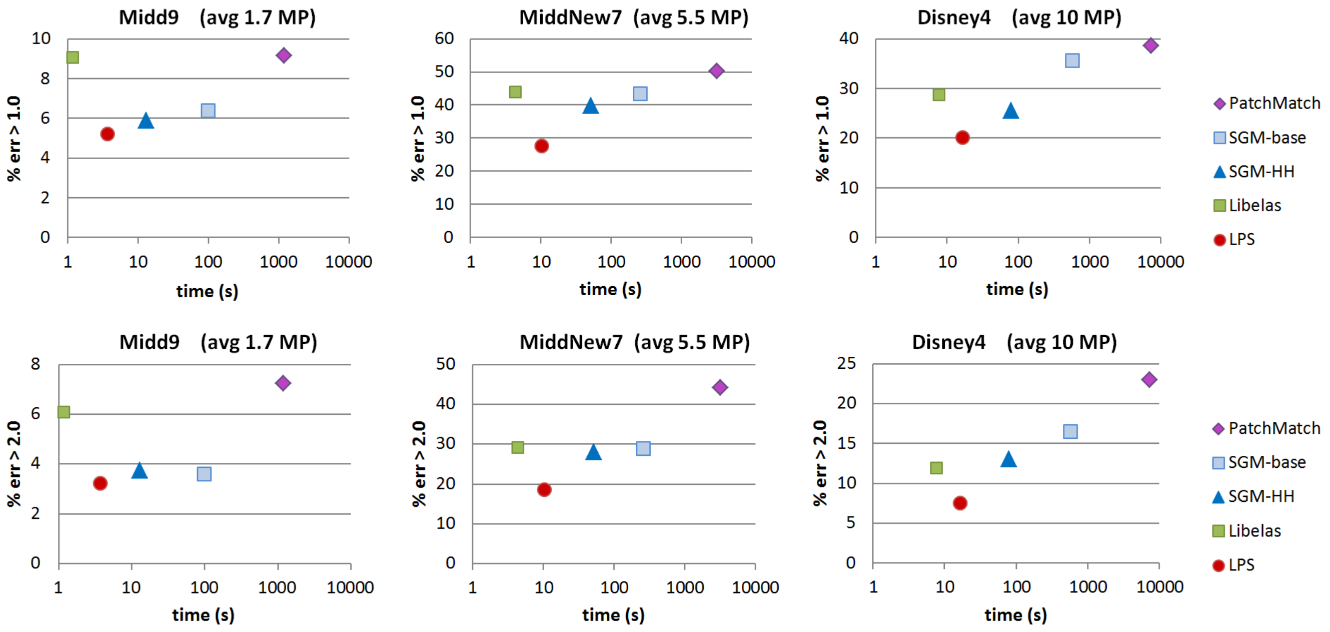

Figure 2. Average error vs. log runtime for error thresholds $t = 1.0$ (top) and $t = 2.0$(bottom) for the three test sets. Our method yields the lowest errors and the second-lowest runtimes.

			PatchMatch			SGM-base			SGM-HH			Libelas			LPS (ours)		
	Image Pair	MP	err1.0	err2.0	time	err1.0	err2.0	time	err1.0	err2.0	time	err1.0	err2.0	time	err1.0	err2.0	time
Midd9	Cloth3	1.4	1.04	0.70	904	1.16	0.71	81.1	1.26	0.82	9.95	2.18	1.67	**0.95**	**0.95**	**0.56**	2.55
	Rocks2	1.4	2.16	1.61	939	1.35	0.85	82.9	**1.33**	0.84	11.7	2.11	1.51	**0.95**	1.34	**0.82**	2.51
	Aloe	1.4	2.96	2.13	928	3.49	1.94	82.2	3.14	1.80	10.6	4.12	2.83	**0.96**	**2.87**	**1.73**	3.21
	Baby3	1.5	7.75	6.44	951	5.57	3.05	84.4	**4.09**	**2.41**	10.7	7.66	3.86	**0.98**	5.36	3.06	3.17
	Dolls	1.5	8.92	5.10	1031	9.46	4.56	89.6	**7.08**	**3.20**	11.6	8.70	4.81	**1.13**	8.52	4.02	3.78
	Art	1.5	6.92	5.45	1060	7.83	5.17	89.0	6.96	4.76	11.5	9.89	7.28	**1.09**	**5.83**	**4.43**	3.22
	Cones	2.7	9.76	6.48	2079	5.47	2.19	160	5.28	2.51	19.7	9.40	4.13	**2.01**	**3.66**	**1.59**	6.12
	Teddy	2.7	14.4	9.83	2065	12.4	**7.44**	157	12.4	8.69	21.3	17.9	12.0	**2.00**	**10.7**	7.55	6.50
	Lamp2	1.4	28.5	27.5	986	10.8	6.61	83.8	11.5	8.63	11.3	19.4	16.6	**0.90**	**5.39**	**3.80**	2.79
MiddNew7	Motorcycle	5.9	33.8	24.2	3330	34.3	16.6	268	29.3	15.1	51.4	34.0	19.1	**4.95**	**12.2**	**6.51**	9.64
	Pipes	5.7	30.9	25.0	3367	29.5	18.7	279	25.3	17.4	54.9	30.7	19.5	**4.53**	**15.6**	**9.80**	9.51
	Adirondack	5.7	73.8	70.3	3212	41.9	26.8	268	43.3	31.0	53.6	42.3	26.9	**4.21**	**29.9**	**18.6**	10.6
	Recycle	5.6	47.6	39.3	3124	42.8	25.3	236	39.0	27.6	51.0	44.2	29.0	**4.28**	**33.6**	**22.2**	10.6
	Piano	5.4	39.0	34.9	3330	35.2	22.4	232	30.0	**20.3**	42.0	40.0	26.5	**3.93**	**28.3**	22.0	11.0
	Playroom	5.3	71.8	67.1	2978	49.9	32.4	283	45.3	30.3	56.2	52.0	33.6	**4.34**	**41.4**	**29.1**	11.0
	Playtable	5.0	55.5	48.4	3186	71.0	59.6	234	65.8	54.5	49.3	63.1	48.7	**4.27**	**33.4**	**24.3**	9.77
Disney4	Couch	10.8	–	–	7821*	–	–	632*	14.4	5.76	85.3	17.7	3.32	**8.68**	**8.06**	**1.43**	14.5
	Mansion	18.9	–	–	13687*	–	–	1107*	16.9	7.19	158	23.3	9.88	**15.9**	**16.6**	**6.98**	24.0
	Statue	4.5	34.3	22.3	3792	24.9	7.70	262	27.5	13.1	36.0	29.5	12.2	**3.02**	**17.7**	**6.46**	13.7
	Bikes	4.7	43.0	23.7	4184	46.3	25.2	272	43.6	26.0	36.6	43.9	22.1	**3.74**	**35.2**	**14.5**	14.2

Table 1. Accuracy (% bad pixels) for for error thresholds $t = 1.0$ and $t = 2.0$ as well as runtimes in seconds for all 20 pairs used in our evaluation. The lowest values in each category are highlighted in bold. *PatchMatch and SGM-base cannot handle the two largest Disney pairs, so the runtimes are extrapolated.

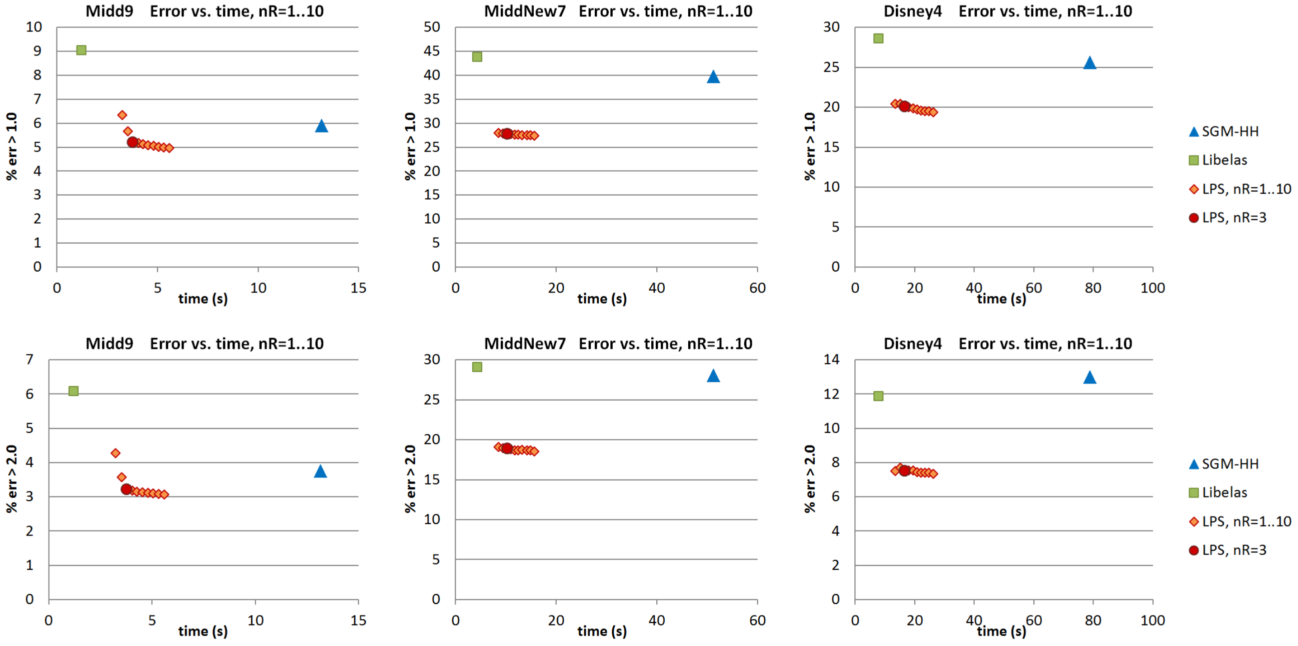

Figure 3. Accuracy vs. runtime of our method for error thresholds $t = 1.0$ (top) and $t = 2.0$ (bottom) for the three test sets as the number of rounds nR is varied from 1 to 10. We use nR=3 for all results reported.

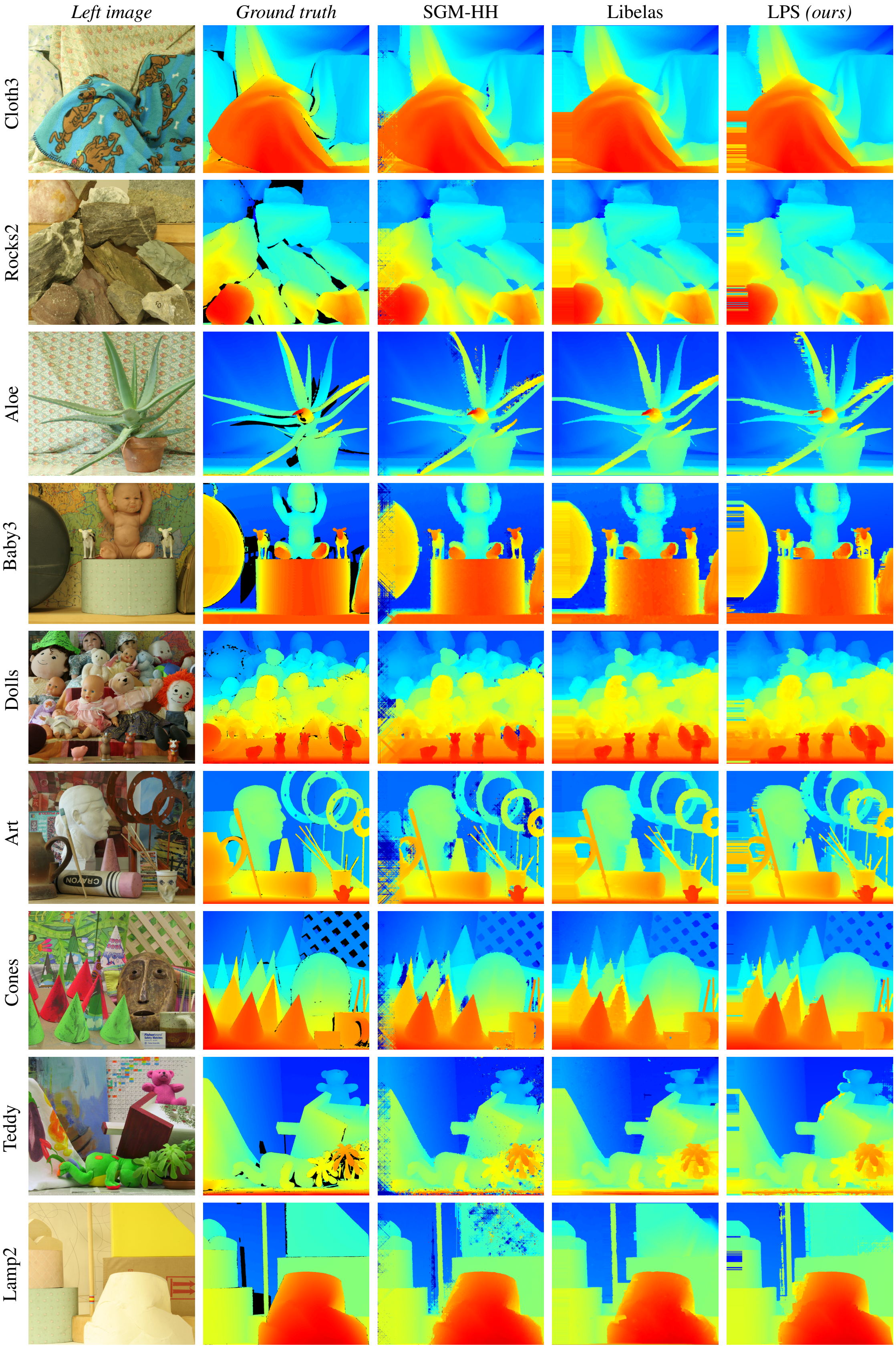

Figure 4. Results by the best three methods on the *Midd9* group, consisting of 9 stereo pairs ranging from 1.4 to 2.7 megapixels, selected from the "full-resolution" 2003–2006 Middlebury datasets. Our LPS method yields highest average accuracy in nonoccluded regions.

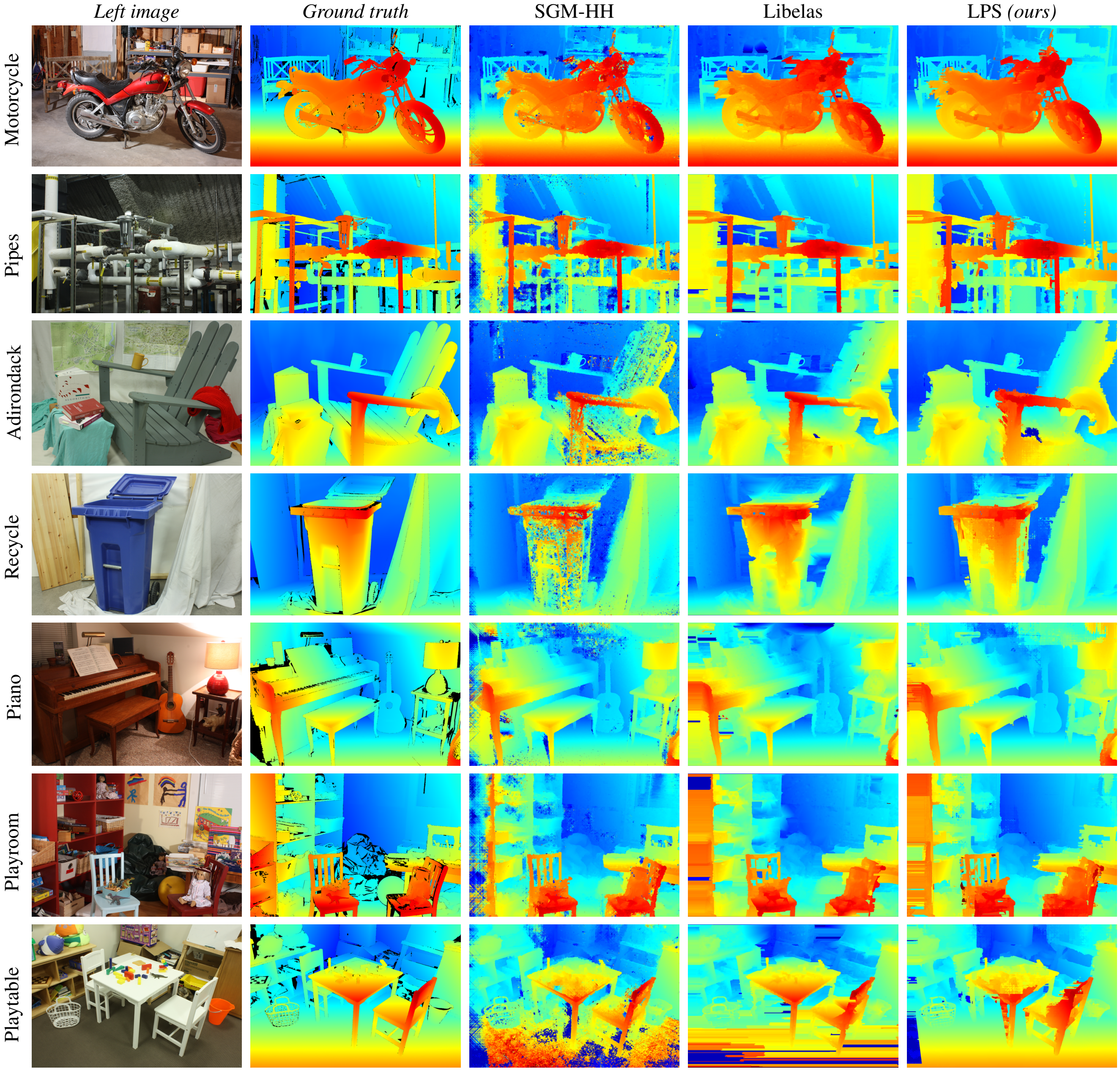

Figure 5. Results by the best three methods on the *MiddNew7* group, consisting of 7 stereo pairs ranging from 5.0 to 5.9 megapixels, selected from the 2014 Middlebury public test dataset. Our LPS method again yields highest accuracy in non-occluded regions.

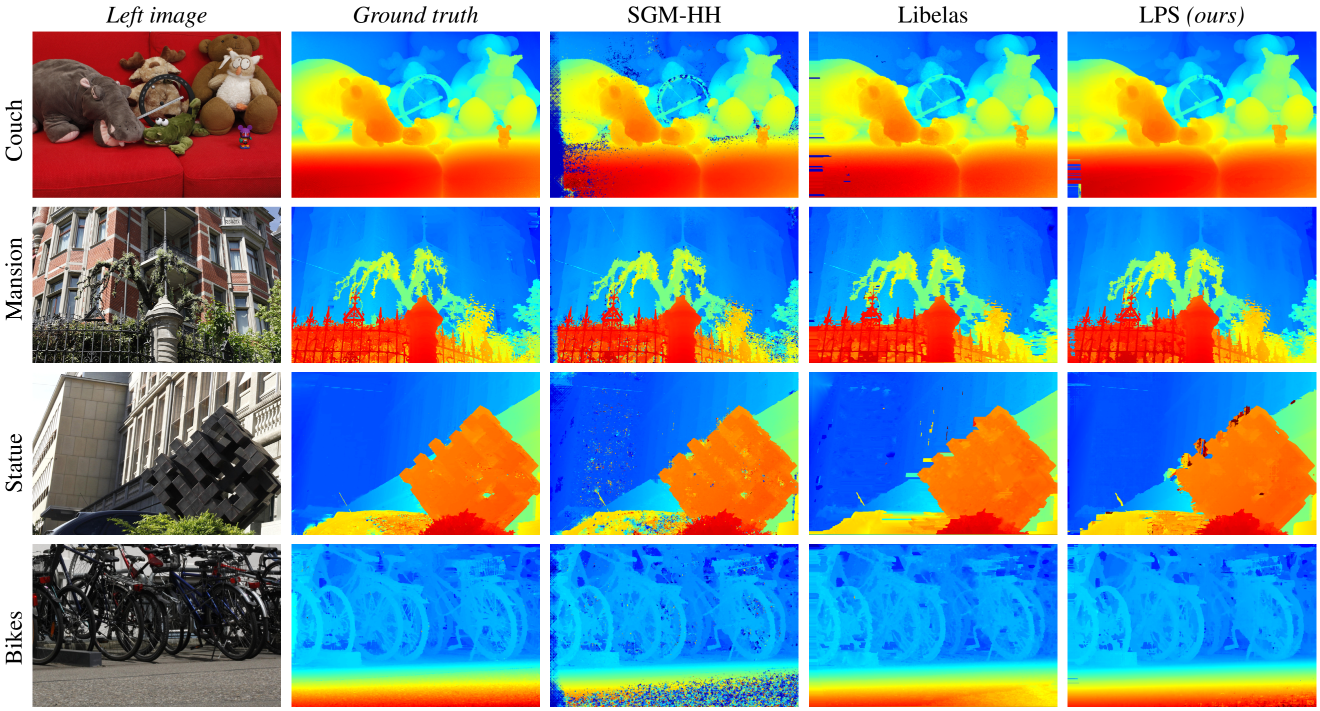

Figure 6. Results by the best three methods on the stereo pairs in the *Disney4* group, ranging from 4.5 to 19 megapixels. Our LPS method yields highest accuracy in non-occluded regions in all cases.

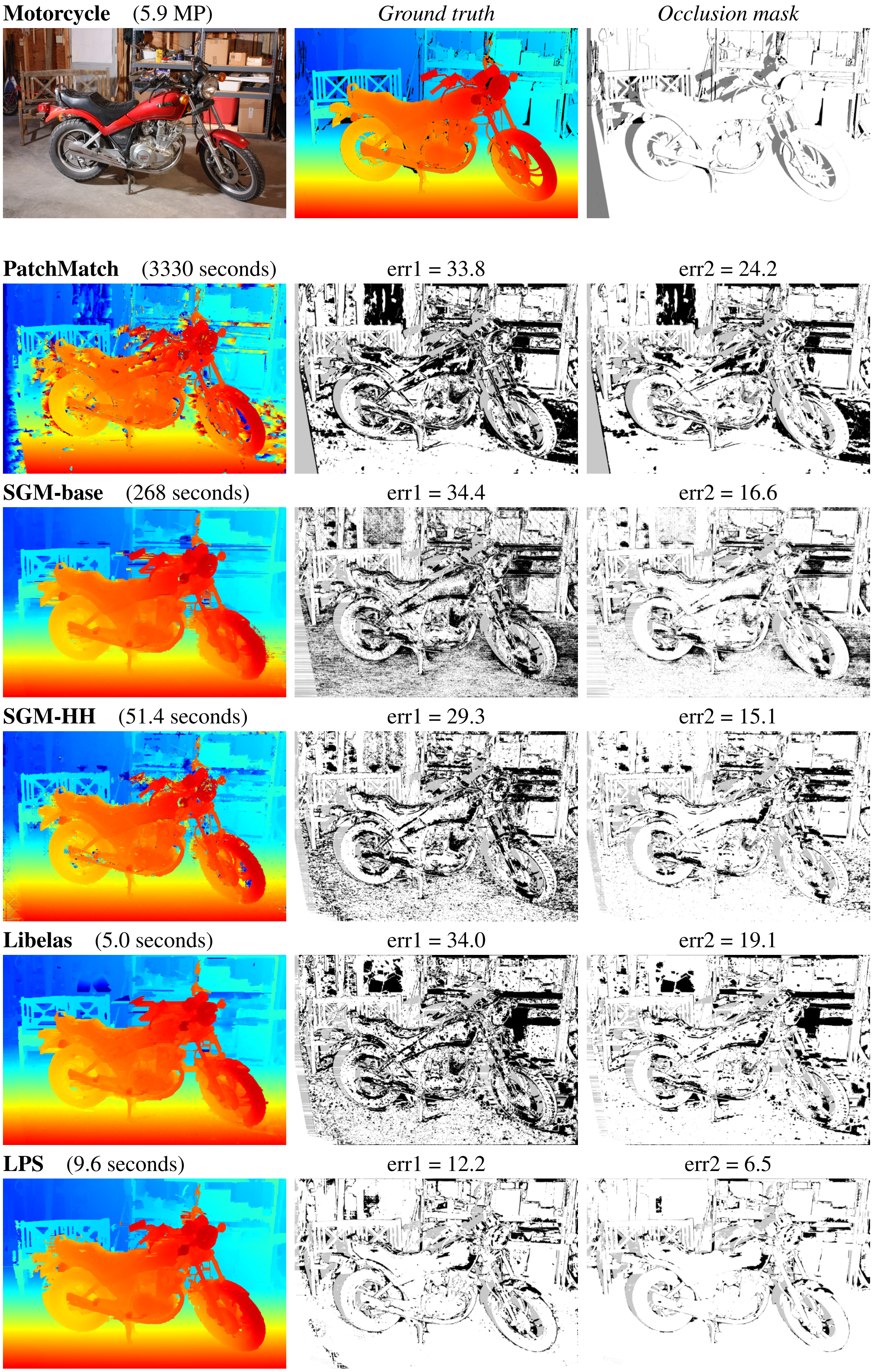

Figure 7. Visualization of errors on the *Motorcycle* pair. The top row shows the left input image, ground truth disparity map, and occlusion mask. The results by the five methods appear below, including disparity maps (left column) and error maps for thresholds $t = 1.0$ (middle column) and $t = 2.0$ (right column). Black pixels in the error maps indicate errors in non-occluded regions. It can be seen that our LPS methods yields significantly fewer errors than the other methods under both thresholds.