MGM: A Significantly More Global Matching for Stereovision – Supplementary Material

Gabriele Facciolo[1]
facciolo@cmla.ens-cachan.fr
Carlo de Franchis[2]
carlo.de-franchis@ens-cachan.fr
Enric Meinhardt[2]
enric.meinhardt@cmla.ens-cachan.fr

[1] IMAGINE/LIGM,
École Nationale des Ponts et Chaussées,
France
[2] CMLA,
École Normale Supérieure de Cachan,
France

1 Algorithm

The pseudocode of the MGM algorithm is shown below (Algorithm 1).

Algorithm 1: MGM

input : pixel-wise cost volume C, search space $\mathcal{D} = \{d_{min}, d_{max}\}$
output : aggregated and over-counting corrected cost volume S_{oc}
$(L_i)_{i=0...7} \longleftarrow 0$ *//initialize L_i, each one is the size of the image $\times|\mathcal{D}|$*
// For each passage
foreach $i \in [0...7]$ **do**
  // traversals as in figure 1
  $(\mathbf{r}, \mathbf{r}^{\perp}) \longleftarrow$ Compute_previous_and_above_directions_for_traversal(i)
  $(x_n)_{n=0...N-1} \longleftarrow$ Compute_adequate_pixel_ordering_for_traversal(i)
  //for all the N pixels of the image
  foreach $n \in [0...N-1]$ **do**
    $\mathbf{p} \longleftarrow x_n$
    $m_{\mathbf{r}} \longleftarrow \min_{d'}(L_i(\mathbf{p}-\mathbf{r}, d'))$ *//precompute the* $\min L_{\mathbf{r}}$
    $m_{\mathbf{r}^{\perp}} \longleftarrow \min_{d'}(L_i(\mathbf{p}-\mathbf{r}^{\perp}, d'))$ *//precompute the* $\min L_{\mathbf{r}^{\perp}}$
    foreach $d \in \mathcal{D}$ **do**
      //this implements equation 9 in the main paper

$$L_i(\mathbf{p},d) \leftarrow C_{\mathbf{p}}(d) + \tfrac{1}{2}\min(L_i(\mathbf{p}-\mathbf{r},d), L_i(\mathbf{p}-\mathbf{r},d\pm 1)+P1, m_{\mathbf{r}}+P2) + \tfrac{1}{2}\min(L_i(\mathbf{p}-\mathbf{r}^{\perp},d), L_i(\mathbf{p}-\mathbf{r}^{\perp},d\pm 1)+P1, m_{\mathbf{r}^{\perp}}+P2)$$

$S_{oc} \longleftarrow \sum_{i\in[0...7]} L_i - 7C$

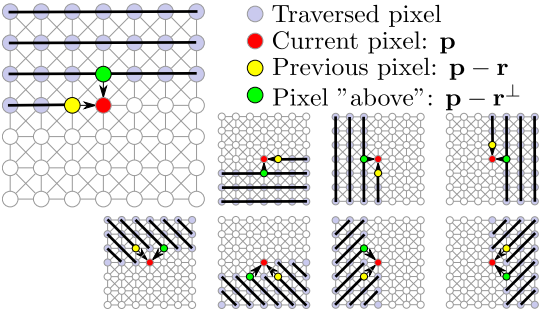
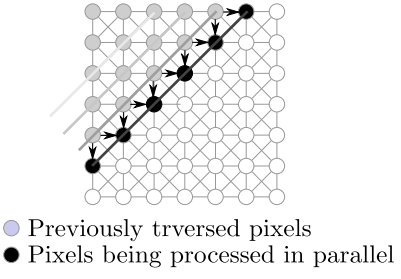


Figure 1: Depiction of the 8 image traversals and the corresponding recursion directions $\mathbf{r}$ and $\mathbf{r}^{\perp}$ in MGM.



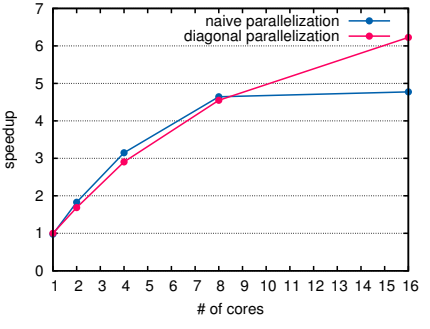


Figure 2: Speedup of MGM applying different parallelization strategies. The naïve strategy consists in computing the 8 image traversals in parallel. This doesn't scale beyond 8 cores. The *diagonal* parallelization (illustrated on the left for the raster ordered image traversal) scales better with the number of cores.

1.1 Parallelization considerations

Since MGM introduces dependency among the scan lines, these cannot be processed in parallel as it is usually done with SGM. A naïve parallelization of MGM consists in performing the 4 or 8 image traversals in parallel. However, the performance improvement with this approach is limited to the number of traversals.

Instead, parallelization in MGM is achieved diagonal-by-diagonal. That is, during the image traversal in raster order (first diagram of figure 1) the pixels of a diagonal going from top-right to bottom-left can be processed in parallel, because they only depend on their top and left neighbors (which are computed by the previous diagonal). This strategy is illustrated in figure 2. A nearly linear speed-up is observed (figure 2) when applying this strategy on a 16-core test machine (Xeon@2.60GHz).

	Tsukuba(16 labels)				Teddy(60 labels)				Venus(20 labels)				Fountain(143 labels)			
Method	**Egap**	**C**	**V**	**C/V**	**Egap**	**C**	**V**	**C/V**	**Egap**	**C**	**V**	**C/V**	**Egap**	**C**	**V**	**C/V**
BP-S	1.78	939	206	4.5	0.68	3295	456	7.2	0.68	2163	223	9.7	1.32	2932	717	4.1
Expansion	0.09	928	199	4.7	0.13	3305	425	7.8	0.07	2171	200	10.8	-0.05	2911	688	4.2
TRW-S	0.00	933	192	4.8	0.00	3293	432	7.6	0.00	2164	205	10.5	0.00	2899	703	4.1
SGM4	48.3	844	826	1.0	21.7	2968	1565	1.9	31.4	2003	1110	1.8	29.0	2611	2036	1.3
ocSGM4	41.9	904	693	1.3	18.2	3225	1179	2.7	25.6	2130	846	2.5	26.5	2794	1763	1.6
MGM4	7.51	918	292	3.1	5.53	3229	703	4.6	4.19	2144	325	6.6	10.7	2779	1207	2.3

Table 1: Energy ratios. **E** is the energy gap ($\frac{E_{ref}-E}{E_{ref}} \times 100$) of the solution with respect to the reference solution of TRW-S (Tree-Reweighted Message Passing [9]), **C/V** is the ratio in the final energy between the data term (**C**) and the regularity term (**V**). All the reported energies are measured in thousands. Two other reference methods are also included: Expansion-move algorithm [2], and BP-S a sequential Belief-Propagation algorithm from [12]. The energy ratios **C/V**, indicate the current balance of the data and regularity term in the solution. In both SGM4 and ocSGM4, the regularity term (**V**) is always much larger than in the solutions of the other methods. This means that the regularity term is weakly enforced by SGM4 and ocSGM4. The solutions of MGM4 on the other hand are much higher and closer to the reference optimization algorithms.

2 Experiments

2.1 In SGM the regularity term is only weakly enforced

As mentioned in the main paper, we observe that in the SGM algorithm the regularity defined in the energy

$$E(D) = \underbrace{\sum_{\mathbf{p}\in I} C_{\mathbf{p}}(D_{\mathbf{p}})}_{\mathbf{C}} + \underbrace{\lambda \sum_{(\mathbf{p},\mathbf{q})\in\mathcal{E}} V(D_{\mathbf{p}}, D_{\mathbf{q}})}_{\mathbf{V}} \tag{1}$$

is only weakly enforced. Whereas for MGM yields a result closer to the reference optimization techniques. To corroborate this, we evaluated the energy of the solutions reported in section 4.1 (table 1) of the main paper. In table 1 we distinguishing the contributions of the data term (**C**) and the regularity term (**V**) for each solution, and compare the ratios $\mathbf{C}/\mathbf{V}$. First, we note that for all the methods the data term **C** have approximately the same energy. However the regularity terms for the results of SGM4 and ocSGM4 are abnormally high, which means that the regularity term is not satisfied by the solution. MGM4, on the other hand, yields much lower values for the regularity term which means that the obtained solutions are more regular.

2.2 MGM systematically improves the energy minima with respect to the baseline SGM

To complement the experiment of section 4.1 we propose evaluate the energy gap for the stereo correspondence problem between the solutions obtained with MGM the baseline SGM with 8 directions (the larger the gap the better). We use for this experiment full resolution stereo-paris from the datasets [8, 10]. The energy is measured on the 8-connected graph, and the smoothness term is

$$V(d,d') = \begin{cases} 0 & \text{if } d = d' \\ P1 & \text{if } |d-d'| = 1 \\ P2 & \text{otherwise} \end{cases}, \tag{2}$$

	Image	$1-\left(\frac{E_{ocSGM}}{E_{SGM}}\right)$	$1-\left(\frac{E_{MGM}}{E_{SGM}}\right)$
GTMid 2014	Motorcycle	0.090	0.419
	Jadeplant	0.059	0.401
	Adirondack	0.050	0.358
	Playtable	0.093	0.482
	Playroom	0.057	0.289
	Teddy	0.142	0.461
	Vintage	0.075	0.523
	Pipes	0.096	0.381
	Shelves	0.058	0.549
	Recycle	0.060	0.515
	Piano	0.092	0.504
GTMid 2006	Baby1	0.126	0.446
	Baby3	0.131	0.449
	Baby2	0.115	0.451
	Plastic	0.046	0.482
	Aloe	0.153	0.311
	Cloth2	0.124	0.298
	Cloth3	0.174	0.271
	Cloth1	0.175	0.269
	Cloth4	0.154	0.311
	Lampshade1	0.084	0.514
	Wood1	0.133	0.511
	Wood2	0.076	0.541
	Lampshade2	0.062	0.520
	Flowerpots	0.071	0.423
	Monopoly	0.091	0.312
	Bowling1	0.073	0.501
	Bowling2	0.090	0.475
	Midd2	0.078	0.518
	Midd1	0.099	0.502
	Rocks1	0.178	0.376
	Rocks2	0.185	0.369
	Books	0.133	0.382
GTMid 2005	Moebius	0.150	0.378
	Reindeer	0.142	0.410
	Dolls	0.124	0.322
	Laundry	0.129	0.371
	Art	0.115	0.374
	AVERAGE	0.107	**0.420**

Table 2: Energy gap **(the larger the better)** between the solutions obtained with the proposed method and the baseline SGM with 8 directions. The images are from the stereo datasets [8, 10]. The energy is measured on the 8-connected graph, in all the test we use $P1 = 8$ and $P2 = 32$ for the smoothness term (2). The energy obtained with MGM is on average 40% lower than SGM. While ocSGM is only 10% lower than SGM on average.

with $P1 = 8$ and $P2 = 32$ for all the experiments.

In table 2 we see that the energy of the solutions obtained with MGM are on average 40% lower than SGM. While ocSGM is only 10% lower than SGM on average.

2.3 Post-processing with median filter

As the evaluation presented in the main paper focuses on energy minimization, no post-processing of the outputs is considered. However, a key component of the SGM algorithm [7] is the 3×3 median filtering of the disparity maps to fill-in small holes. Here we compare the impact of this post-process (denoted +med) on the baseline SGM and MGM. Figure 3 show that the results of SGM improve substantially after median filtering, while the results of MGM do not change much. Nevertheless, even after filtering the results of SGM present more errors than MGM specially in poorly textured areas (as shown in figure 4).

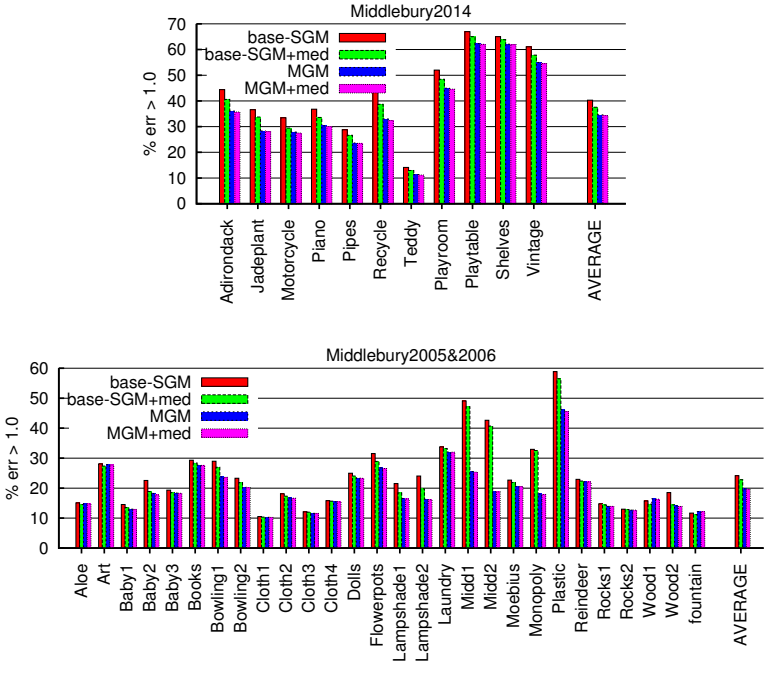


Figure 3: Bad pixel ratios (% of pixels with error > 1) on the Middlebury 2014 and the Middlebury 2005&2006 datasets. The statistics corresponds to results of the baseline SMG and MGM, with and without the post-processing using the 3×3 median filter (+med).

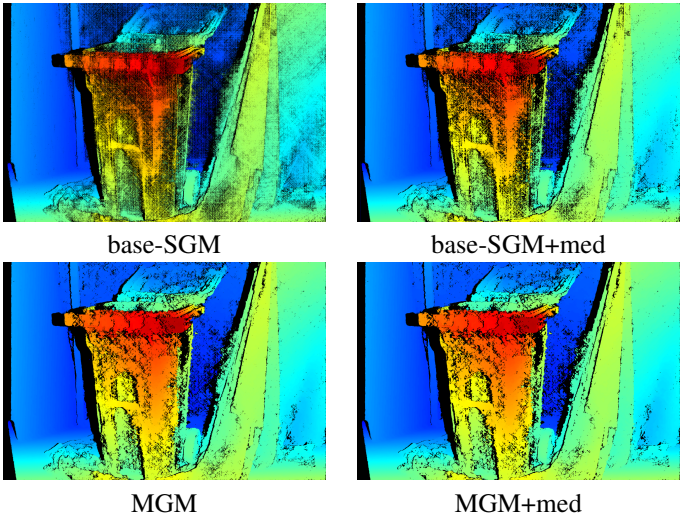


Figure 4: One result for comparing the impact of the median filter (+med) post-process when applied to the baseline SGM and MGM.

3 More results

3.1 Middlebury 2014 dataset [10]

Figures 6 and 7 show more results of our method. We compare the results of our baseline implementation of SGM, our method (MGM), and the publicly available ELAS [4]. The two figures show the sparse and dense results of each method. SGM and MGM use the same settings for all the images: 8 propagation directions and the Hamming distance of census transform [13, 14] on a 5×5 neighborhood (normalized by the number of channels as matching cost $C_{\mathbf{p}}$. The parameters $P1$ and $P2$ were set for all images to $P1 = 8$ and $P2 = 32$. No intensity cues (adaptation of $P2$) were used but they could easily be incorporated [1, 7] in both SGM and MGM. To prevent influence of the post-processing steps on the evaluation, the results shown for base-SGM and MGM in figure 6 correspond to the unaltered outputs of the winner-take-all stage. The bad pixel ratios for the dense results are compared in table 5. The sparse results (figure 7) are obtained applying the left-to-right consistency check [3] with threshold set to 1.

3.2 KITTI dataset [5]

Results on the KITTI training dataset, are shown in table 3. We didn't optimize the parameters, instead we used the ones proposed in [11]: $P1 = 7$, $P2 = 100$. We used the census distance (5×5 windows) as cost, refined the disparities with V-fit sub-pixel interpolation [6], then filtered with a 3 median filter and with the left-to-right consistency check. The running time for computing each frame on a 16-core computer is about 6 seconds for MGM, and 5.3 seconds for base-SGM. The results shown in figure 8 confirm that MGM yields slightly denser results than base-SGM. Evaluating the MGM algorithm on the test database yields table 4.

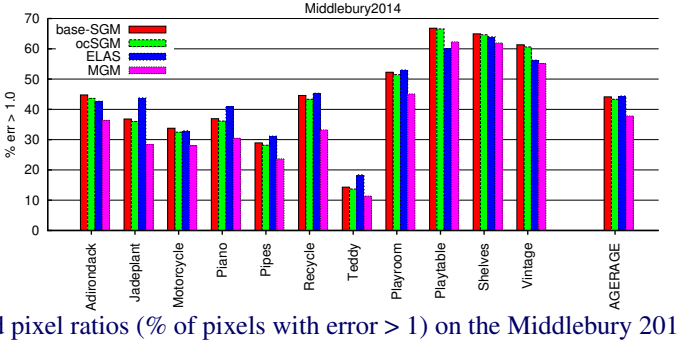


Figure 5: Bad pixel ratios (% of pixels with error > 1) on the Middlebury 2014 test images.

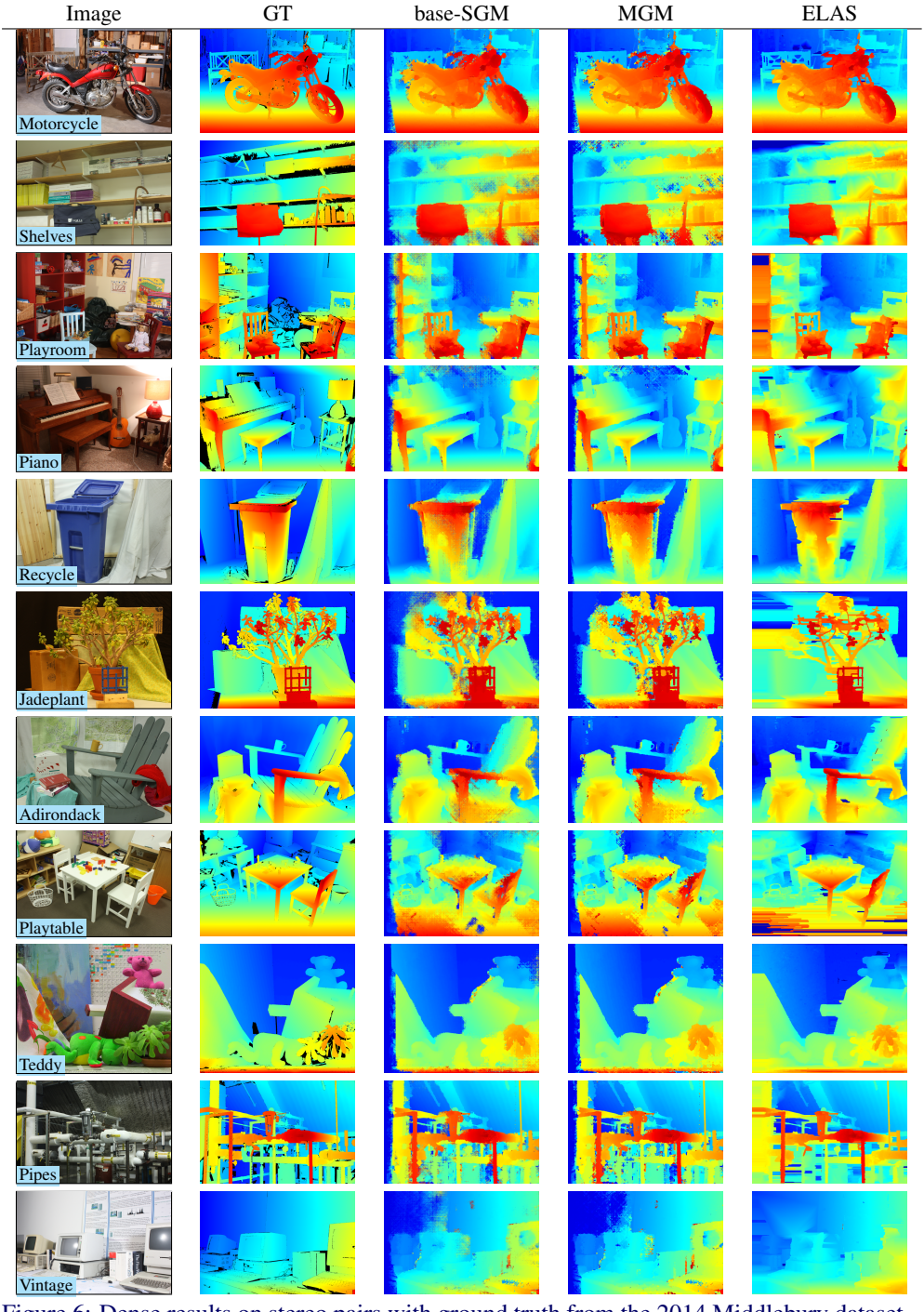


Figure 6: Dense results on stereo pairs with ground truth from the 2014 Middlebury dataset. We compare the results of our baseline implementation of SGM, MGM, and ELAS [4].

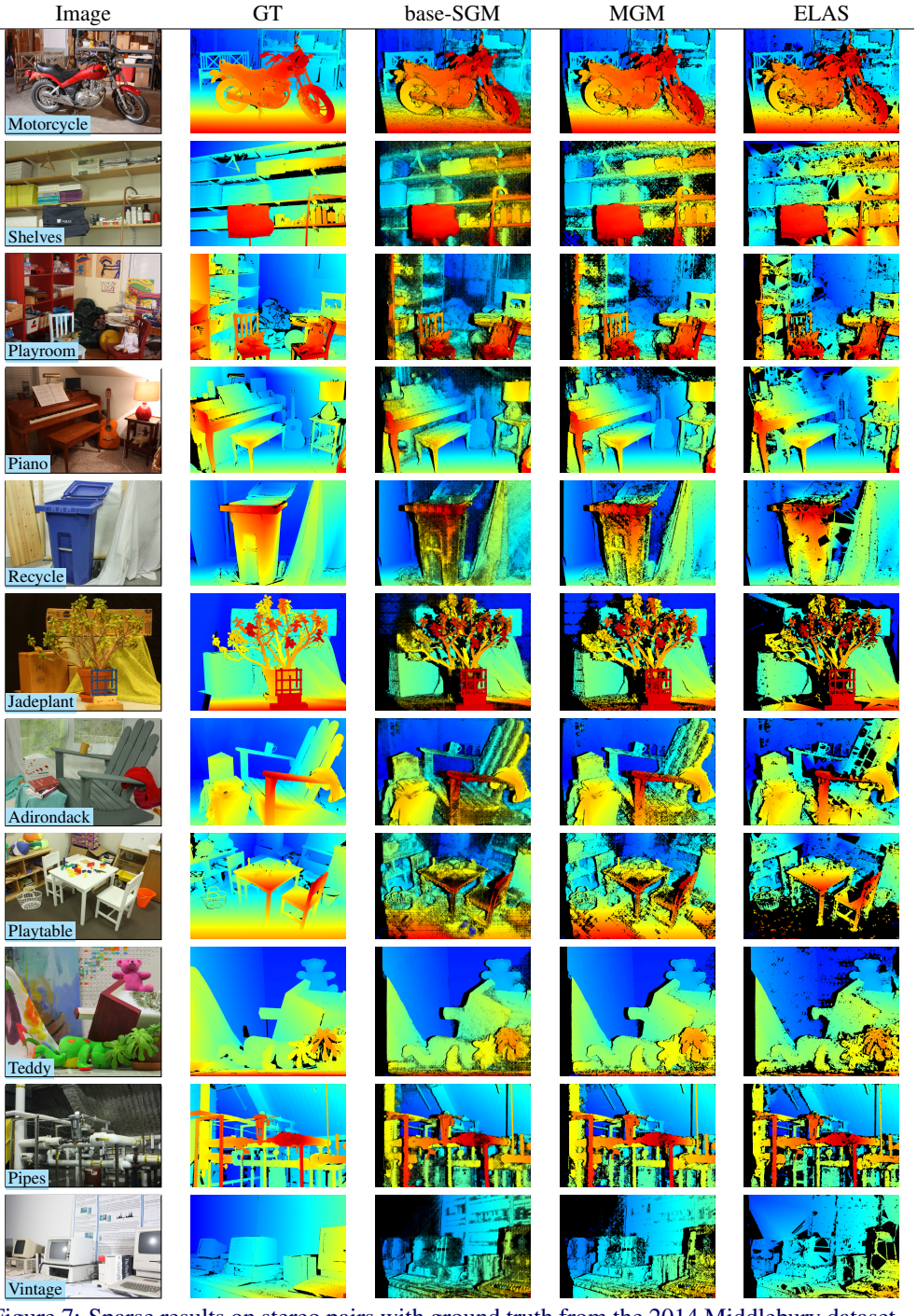
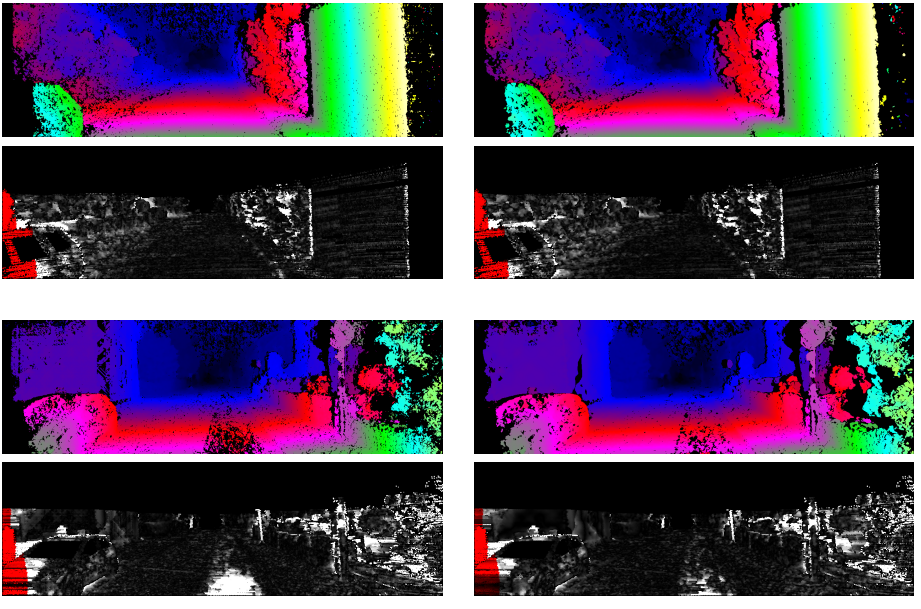


Figure 7: Sparse results on stereo pairs with ground truth from the 2014 Middlebury dataset. We compare the results of our baseline implementation of SGM, MGM, and ELAS [8].

method	Out-Noc (all px)	Out-Noc (estimated px)	Density
base-SGM	7.30 %	**3.94 %**	88.78 %
MGM	**6.33** %	4.09 %	**92.18**%

Table 3: Average results of the MGM and base-SGM algorithms on the KITTI **training** dataset. Out-Noc is the fractionof erroneous pixels (computed using the threshold at 3 pixels) in non-occluded areas, evaluated on all the pixels (interpolated), or restricted to the estimated ones. The average densities of the disparity maps are shown in the last column.



base-SGM MGM

Figure 8: **Results of base-SGM and MGM for two stereo pairs from training dataset of the KITTI challenge.**

Error	Out-Noc	Out-All	Avg-Noc	Avg-All
2 pixels	8.51 %	9.95 %	1.2 px	1.4 px
3 pixels	5.70 %	6.89 %	1.2 px	1.4 px
4 pixels	4.31 %	5.30 %	1.2 px	1.4 px
5 pixels	3.45 %	4.28 %	1.2 px	1.4 px

Table 4: Results of the MGM algorithm on the KITTI **testing** dataset. We used the census distance (5×5 windows), with parameters $P1 = 7$, $P2 = 100$ (as in [11]) the disparities are refined with V-fit sub-pixel interpolation [6], then filtered with a 3 median filter, and with the left-to-right consistency check. The running time for each image is about 6 seconds on a 16-core computer.

References

[1] M. Bleyer and M. Gelautz. Simple but effective tree structures for dynamic programming-based stereo matching. In *Proc. VISAPP*, pages 415–422, 2008.

[2] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Trans. Pattern Anal. Machine Intell.*, 23(11):1222–1239, 2001.

[3] P. Fua. A parallel stereo algorithm that produces dense depth maps and preserves image features. *Machine Vision and Applications*, 6(1):35–49, December 1993.

[4] A. Geiger, M. Roser, and R. Urtasun. Efficient large-scale stereo matching. In *Lecture Notes in Comput. Sci.*, volume 6492 LNCS, pages 25–38, 2011.

[5] A. Geiger, P. Lenz, and R. Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In *Proc. CVPR*, 2012.

[6] Istvan Haller and Sergiu Nedevschi. Design of interpolation functions for subpixel-accuracy stereo-vision systems. *IEEE Trans. on Image Proc.*, 21(2):889–898, 2012.

[7] H. Hirschmüller. Stereo processing by semiglobal matching and mutual information. *IEEE Trans. Pattern Anal. Machine Intell.*, 30(2):328–41, February 2008.

[8] H. Hirschmüller and D. Scharstein. Evaluation of stereo matching costs on images with radiometric differences. *IEEE Trans. Pattern Anal. Machine Intell.*, 31(9):1582–1599, September 2009.

[9] V. Kolmogorov. Convergent tree-reweighted message passing for energy minimization. *IEEE Trans. Pattern Anal. Machine Intell.*, 28(10):1568–1583, 2006.

[10] D. Scharstein, H. Hirschmüller, Y. Kitajima, K. Krathwohl, N. Nešić, X. Wang, and P. Westling. High-resolution stereo datasets with subpixel-accurate ground truth. In *Pattern Recognition*, pages 31–42. Springer, 2014.

[11] R. Spangenberg, T. Langner, S. Adfeldt, and R. Rojas. Large scale semi-global matching on the cpu. In *Proc. IEEE Intelligent Vehicles Symposium*, pages 195–201, June 2014.

[12] R. Szeliski, R. Zabih, D. Scharstein, O. Veksler, V. Kolmogorov, A. Agarwala, M. Tappen, and C. Rother. A comparative study of energy minimization methods for Markov random fields with smoothness-based priors. *IEEE Trans. Pattern Anal. Machine Intell.*, 30(6):1068–1080, 2008.

[13] R. Zabih and J. Woodfill. Non-parametric local transforms for computing visual correspondence. *Proc. ECCV*, 2:151–158, 1994.

[14] K. Zhu, P. D'Angelo, and M. Butenuth. Evaluation of stereo matching costs on close range, aerial and satellite images. In *Proc. ICPRAM*, pages 379–385. SciTePress, 2012.