Efficient and Accurate Depth Estimation with 1-D Max-Tree Matching

Rafaël Brandt[1] **Nicola Strisciuglio**[1] **Nicolai Petkov**[1]

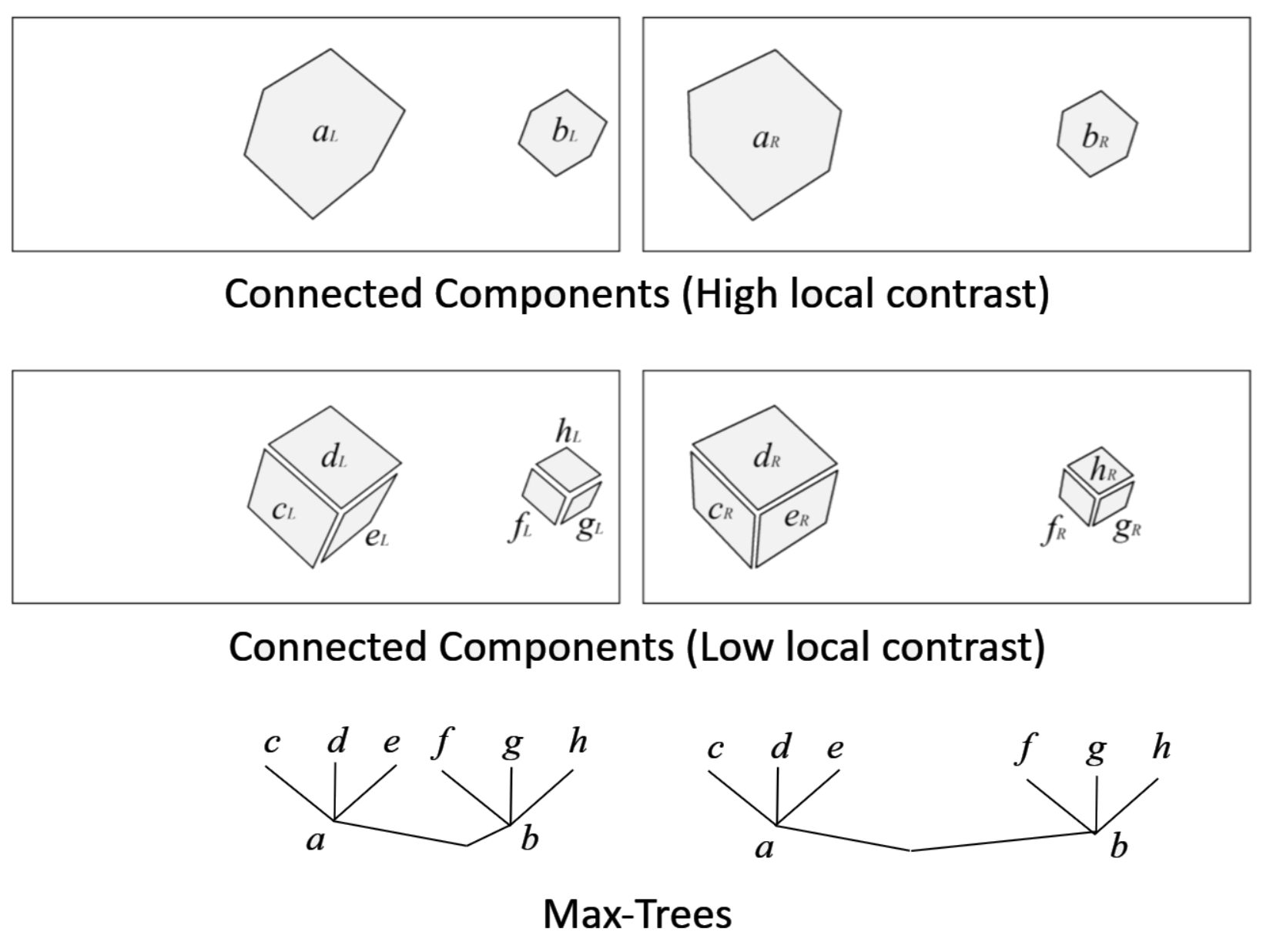
[1]Bernoulli Institute, University of Groningen, P.O. Box 407, 9700 AK Groningen, The Netherlands

Introduction

In stereo matching, the three-dimensional structure of a scene is recovered by finding corresponding pixels in image pairs. The similarity of two pixels is quantitatively computed by a matching cost function. For a pixel in the reference image, the match-able pixel in the second image with the lowest matching cost is selected. Efficient yet accurate extraction of depth from stereo image pairs is required by several systems with low power resources, such as robotics and embedded systems. CNN-based methods compute highly accurate disparity maps, but their need of power-consuming GPUs to compute the many convolutions they are composed of limit their usability on embedded or power-constrained systems. When GPUs cannot be easily used, algorithms for depth estimation are required to provide a reasonable trade-off between accuracy and computational efficiency.

Method

Objects can be recursively decomposed into sets of smaller objects. The MTStereo algorithm that we present exploits contrast information of objects in a hierarchical fashion for efficient stereo matching.



Method (continued)

Our method performs the following steps:

Pre-processing: We detect edges in a rectified input image pair by convolution with a Sobel kernel. Subsequently, we perform color quantization to obtain shallower trees which are less expensive to match.

Max-Tree construction: The Max-Tree allows storing the hierarchy of connected components resulting from different thresholds. We compute 1D Max-Trees based on the pre-processed image pair, i.e. one based on each row. This can be efficiently constructed with a single pass on the image.

Coarse-to-fine matching: We initially match coarse nodes in the tree structures. This can be done efficiently since they are small in number. Thereafter, only descendants of already matched nodes are matched. We deployed a cost function that takes into account contextual information on the Max-trees and a cost aggregation method that preserves disparity edges.

Refinement: Occluded and outlier nodes are removed.

Reliable node extrapolation: The disparity map is made more dense and accurate by computing median disparity values of nodes which are vertical neighbors. A disparity map is computed through linear interpolation of endpoint disparity of the matched nodes with the lowest coarseness level.

Guided pixel matching: The resulting disparity map was generated assuming all surfaces are perfectly flat due to the used linear interpolation of endpoint disparities. Guided pixel matching is performed to recover surface shape.

Refinement: Occluded and outlier pixels are removed.

Qualitative Results

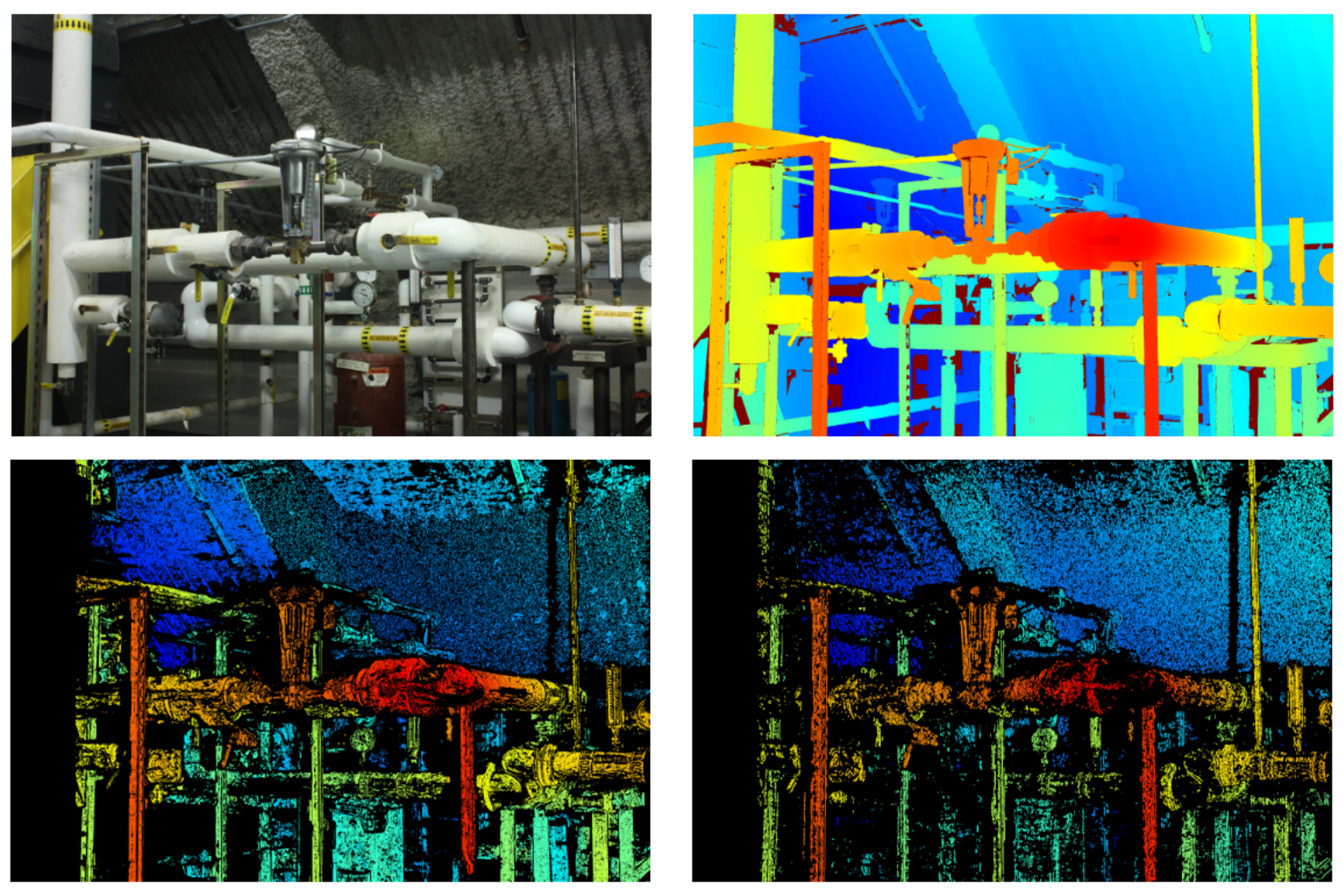


Figure: Reference image (top left), ground truth (top right), our semi-dense estimation (bottom left), our sparse estimation (bottom right).

Conclusions

We proposed a stereo matching method, called MT-Stereo, for systems with low power resources which require efficient and accurate depth estimation. It is based on a hierarchical representation of image pairs with Max-Trees, which we use to identify matching regions along image scan-lines. We deployed a cost function that takes into account contextual information on the Max-trees and a cost aggregation method that preserves disparity edges. Although computing sparse disparity maps, the MTStereo algorithm achieves outputs dense enough for many applications of robot navigation or visual servoing. It does not require GPU computations and can run on devices with low power availability.

The code is available at `https://github.com/rbrandt1/MaxTreeS/`.

Overview of Intermediate Stages

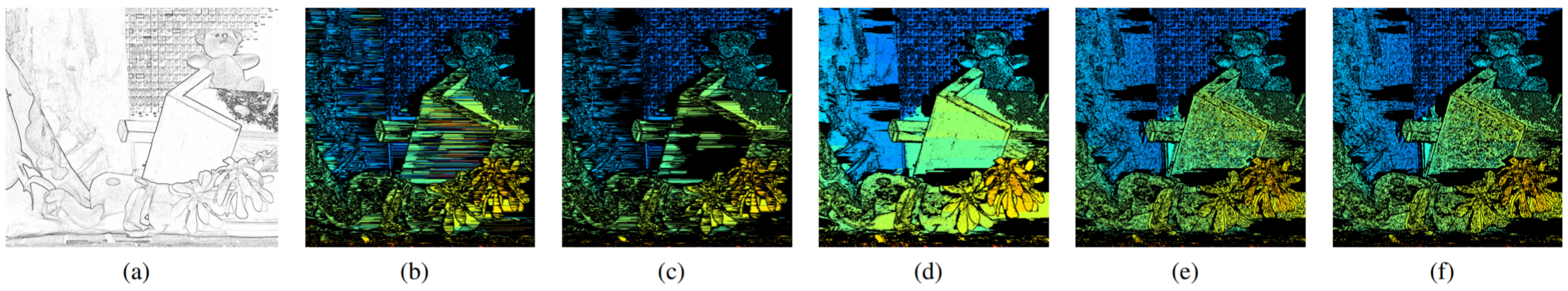


Figure: (a) Pre-processed image. (b) Coarse-to-fine matching. (c) Refinement. (d) Reliable node extrapolation. (e) Guided pixel matching. (f) Refinement.

Acknowledgements

This work received support from the EU Horizon 2020 program, under the project TrimBot2020 (grant No. 688007)

Quantitative Results

Table: The average absolute disparity error (in pixels) among all pixels for which a disparity was estimated and the corresponding ground truth values on the Middlebury benchmark. Our results are rendered bold.

MotionStereo [1]	**SP**	SNCC [2]	LS-ELAS [3]	ELAS [4]	SED [5]	**SD**	SGBM1 [6]	SGBM2 [6]
1.72	2.35	3.25	4.35	4.94	5.38	6.51	7.83	8.92

Table: Time (in seconds) to process image pairs of 1 MP (Intel® Core^{TM} i7-2600K CPU @3.4GHz).

	Middlebury [7]		Kitti2015 [8]	Real Garden [9]	Synth Garden [10]	Driving [11]	Monkaa [11]	Flying3D [11]
	time/MP		time/MP	time/MP	time/MP	time/MP	time/MP	time/MP
SP	2.41	23.74	3.58	2.65	2.90	3.60	2.84	3.47
SD	2.72	26.92	4.54	2.87	2.94	4.29	2.97	3.86

References

- J. Valentin, A. Kowdle, J. T. Barron, N. Wadhwa, M. Dzitsiuk, M. Schoenberg, V. Verma, A. Csaszar, E. Turner, I. Dryanovski, *et al.*, "Depth from motion for smartphone ar," in *SIGGRAPH Asia*, p. 193, ACM, 2018.
- N. Einecke and J. Eggert, "A two-stage correlation method for stereoscopic depth estimation," in *DICTA*, pp. 227–234, IEEE, 2010.
- R. A. Jellal, M. Lange, B. Wassermann, A. Schilling, and A. Zell, "Ls-elas: Line segment based efficient large scale stereo matching," in *IEEE ICRA*, pp. 146–152, IEEE, 2017.
- A. Geiger, M. Roser, and R. Urtasun, "Efficient large-scale stereo matching," in *Asian conference on computer vision*, pp. 25–38, Springer, 2010.
- D. Peña and A. Sutherland, "Disparity estimation by simultaneous edge drawing," in *ACCV 2016 Workshops*, pp. 124–135, 2017.
- H. Hirschmuller, "Stereo processing by semiglobal matching and mutual information," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, pp. 328–341, Feb. 2008.
- H. Hirschmuller and D. Scharstein, "Evaluation of cost functions for stereo matching," in *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8, IEEE, 2007.
- M. Menze, C. Heipke, and A. Geiger, "Joint 3d estimation of vehicles and scene flow," in *ISPRS Workshop on Image Sequence Analysis (ISA)*, 2015.
- T. Sattler, R. Tylecek, T. Brox, M. Pollefeys, and R. B. Fisher, "3d reconstruction meets semantics–reconstruction challenge 2017," in *ICCV Workshop, Venice, Italy, Tech. Rep*, 2017.
- R. Tylecek, T. Sattler, H.-A. Le, T. Brox, M. Pollefeys, R. B. Fisher, and T. Gevers, "The second workshop on 3d reconstruction meets semantics: Challenge results discussion," in *ECCV 2018 Workshops*, pp. 631–644, 2019.
- N. Mayer, E. Ilg, P. Häusser, P. Fischer, D. Cremers, A. Dosovitskiy, and T. Brox, "A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation," in *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.