Filtering and Interpolation of Inaccurate and Incomplete Depth Maps

Julia Navarro, Joan Duran and Antoni Buades

Dept. Mathematics and Computer Science, IAC3. Universitat de les Illes Balears, E-07122, Palma, Spain

(julia.navarro@uib.es, joan.duran@uib.es, toni.buades@uib.es)

1. Introduction

Local algorithms for stereo fail to match accurately points at areas not discriminative enough, mainly texture-less regions. For this reason, these approaches usually involve a validation process which discards unreliable estimates. We propose a method for **filtering incorrect depth estimates and filling in those areas which have not been matched**. The proposed model combines total variation regularization with a non local term taking advantage of image self similarity. The method can be easily generalized to deal with other tasks, as for example stereo upsampling.

2. Proposed model

Let $\mathbf{I}_1, \mathbf{I}_2 \in \mathbb{R}^{N\times C}$ be the stereo pair, with N and C being the number of pixels and color channels, respectively. The initial disparity map is denoted by $f \in \mathbb{R}^N$ and the sought solution by $u \in \mathbb{R}^N$. The binary mask indicating match reliability is considered as a vectorized matrix $m \in \{0,1\}^N$.

To increase the density of f and get an improved disparity map u, we present a **variational interpolation approach**:

$$\arg\min_{u\in\mathbb{R}^N} \frac{1}{2}\|u-f\|_{2,m}^2 + \lambda\|\nabla u\|_1 + \mu\|\nabla_\omega u\|_1, \tag{1}$$

where $\lambda, \mu > 0$ are trade-off parameters and $\|\cdot\|_{2,m}$ is the weighted squared Euclidean norm

$$\|u-f\|_{2,m}^2 = \sum_{i=1}^{N} m_i\,(u_i - f_i)^2 .$$

The nonlocal gradient $\nabla_\omega u \in \mathbb{R}^{N\times N}$ is given in terms of the weighted differences

$$(\nabla_\omega u)_{i,j} = \sqrt{\omega_{i,j}}\,(u_j - u_i),$$

where $\omega_{i,j}$ measures the *similarity* between i and j, taking into account spatial closeness and color similarity in image $\mathbf{I}_1$:

$$\omega_{i,j} = \frac{1}{\Gamma_i}\exp\left(-\frac{\|i-j\|^2}{h_{\text{spt}}^2} - \frac{\|\mathbf{I}_1(P_i)-\mathbf{I}_1(P_j)\|^2}{h_{\text{sim}}^2}\right)$$

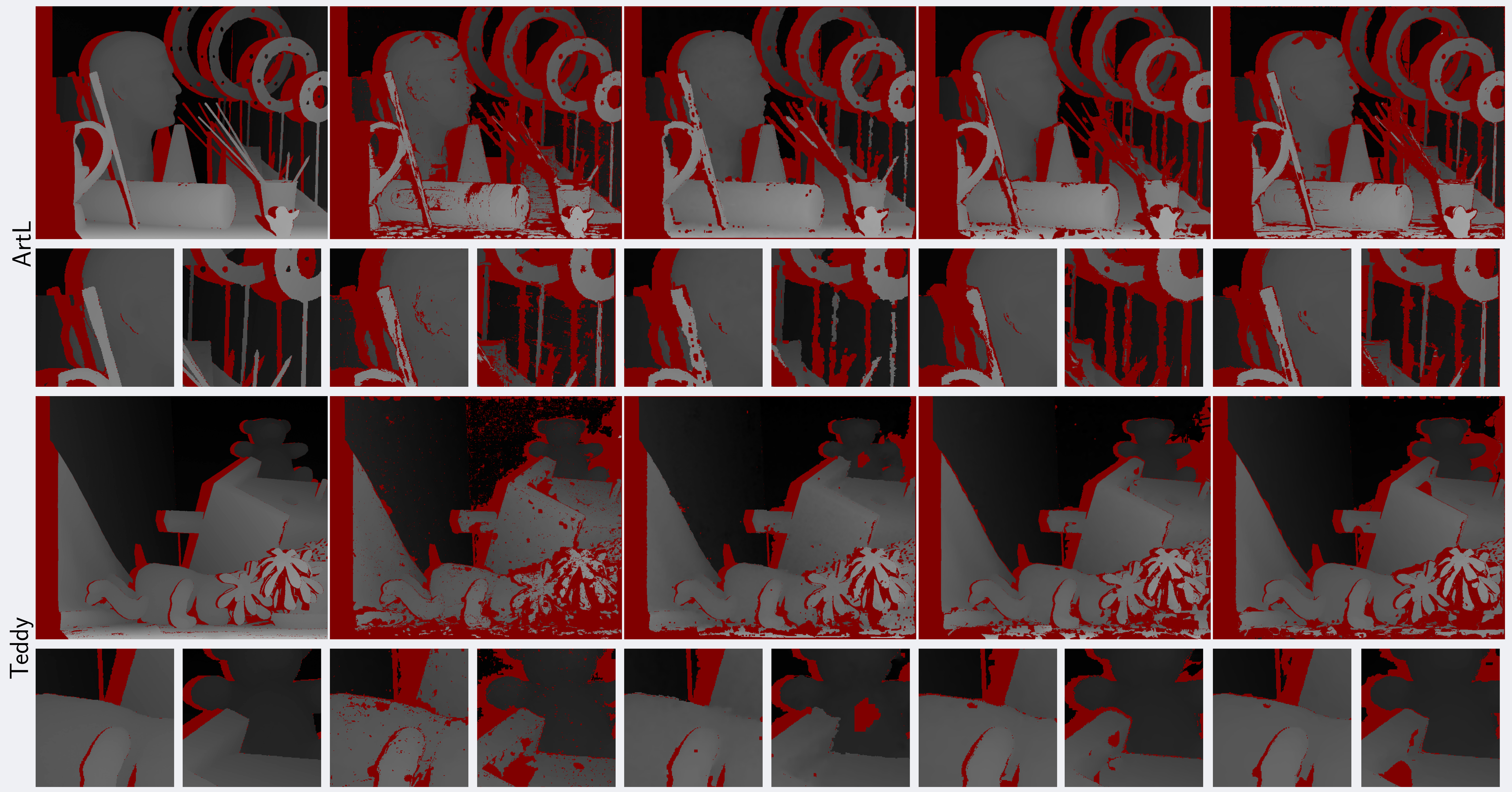
if $\|i-j\|_\infty \le \nu$ and zero otherwise. In this setting, $\nu \in \mathbb{Z}^+$ determines the size of the window to search for similar pixels, P_i denotes a patch centered at pixel i, $\|\cdot\|$ refers to the Euclidean norm, and Γ_i is a normalization factor. The filtering parameters $h_{\text{spt}}, h_{\text{sim}} > 0$ measure the influence of each term.

The minimization problem (1) is convex but non smooth. To find a fast, global optimal solution we use the first-order primal-dual algorithm [1].

3. Experimental results

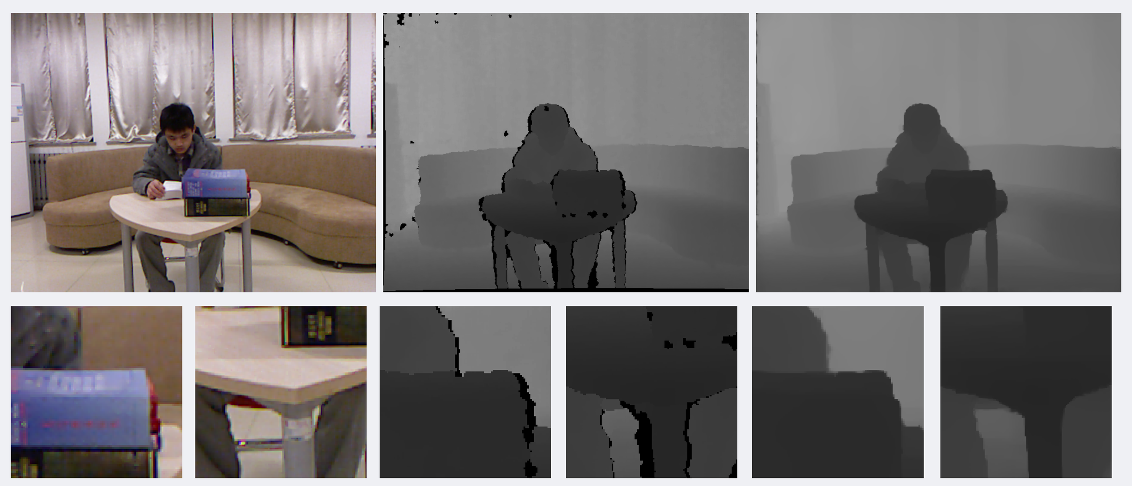
- **Experiments on Middlebury dataset [2]:**

method	density	bad 0.5	bad 1.0	bad 2.0	bad 4.0	avgErr
DAWA [3]	65.13	28.67	10.92	3.78	1.46	0.96
DAWA + ELAS [4]	89.42	54.42	31.36	16.36	8.36	2.83
DAWA + EpicFlow [5]	80.91	35.26	16.45	7.85	3.75	1.49
DAWA + proposed	**73.24**	**32.70**	**14.16**	**5.80**	**2.41**	**1.15**



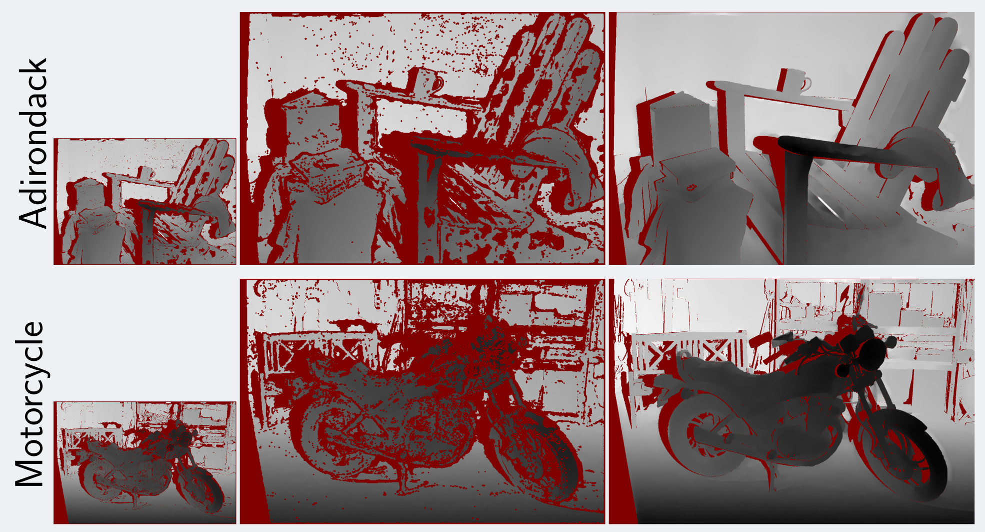
Comparison with state-of-the-art densification methods ELAS and EpicFlow. The initial disparity is the map estimated by the local stereo method DAWA. Red pixels in the ground truth denote occluded points, while they indicate invalidated matches in all the other images.

- **Filtering a depth map from a Kinect:**



Interpolation and filtering of the depth map extracted with a Kinect RGBD camera. Data from [6].

- **Depth upsampling:**



Examples on 2× disparity upsampling. The refinement and interpolation process is applied on the bilinearly upsampled $\hat{u}^H$, resulting on the disparity map u^H. Red pixels in u^H correspond to ground truth occlusions.

4. Conclusions

We have presented a variational interpolation method for incomplete disparity or depth maps. The proposed energy combines local and non local regularization with a fidelity term to the initial map at known areas.

Experiments on the Middlebury stereo database have demonstrated that the method is competitive with respect to the state of the art. Finally, an application to depth upsampling has also been presented.

5. References

[1] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *J. Math. Imaging Vis.*, vol. 40, no. 1, pp. 120–145, 2011.

[2] D. Scharstein, H. Hirschmüller, Y. Kitajima, G. Krathwohl, N. Nešić, X. Wang, and P. Westling, "High-resolution stereo datasets with subpixel-accurate ground truth," in *Proc. German Conference on Pattern Recognition (GCPR)*, 2014, vol. 8753 of *LNCS*, pp. 31–42.

[3] J. Navarro and A. Buades, "Disparity adapted weighted aggregation for local stereo," in *Proc. IEEE International Conference on Image Processing (ICIP)*, 2017, vol. 1, pp. 2249–2253.

[4] A. Geiger, M. Roser, and R. Urtasun, "Efficient large-scale stereo matching," in *Proc. Asian Conference on Computer Vision (ACCV)*, 2010, vol. 6492 of *LNCS*, pp. 25–38.

[5] J. Revaud, P. Weinzaepfel, Z. Harchaoui, and C. Schmid, "EpicFlow: Edge-preserving interpolation of correspondences for optical flow," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 1164–1172.

[6] J. Yang, X. Ye, K. Li, C. Hou, and Y. Wang, "Color-guided depth recovery from rgb-d data using an adaptive autoregressive model," *IEEE Trans. Image Process.*, vol. 23, no. 8, pp. 3443–3458, 2014.

Acknowledgements

The authors were supported by Ministerio de Economía y Competitividad under grants TIN2014-53772-R and TIN2017-85572-P.