ATGV-Net: Accurate Depth Super-Resolution

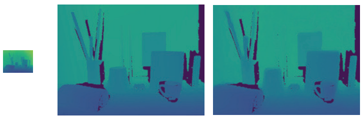
Gernot Riegler   Matthias Rüther   Horst Bischof

Institute for Computer Graphics and Vision, Graz University of Technology, Austria

{riegler, ruether, bischof}@icg.tugraz.at

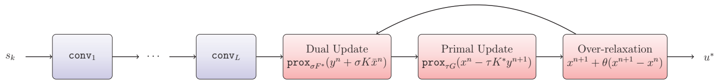
Motivation and Contribution

- Time-of-Flight sensors produce depth maps that are usually of low-resolution and contain severe noise.



Input   GT   Estimate

- For color images, machine learning based methods for single image super-resolution are highly successful [6, 11, 16].
  - Training data is easy to obtain.
- For depth data, methods are mostly based on energy minimization [5, 7].
  - Piece-wise affine surfaces.
  - Sharp depth discontinuities.



- Unroll variational method with anisotropic Total Generalized Variation [2] regularization on top of a deep network.
- Train method end-to-end on huge set of synthetically generated depth maps.
- Code and data is published on GitHub
  `https://github.com/griegler/primal-dual-networks`

ATGV-Net

- Formulated as bi-level optimization problem [14, 15]:

$$\min_{w} \frac{1}{K} \sum_{k=1}^{K} L(u^*(f(w, s_k)), t_k) \qquad \text{(HL)}$$

$$\text{s.t.} \quad u^*(f(w, s_k)) = \arg\min_{u} E(u; f(w, s_k)). \qquad \text{(LL)}$$

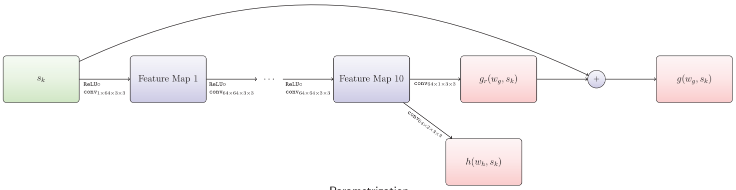
- Energy functional with anisotropic TGV as in [7]
  - Emphasizes piece-wise affine surfaces.
  - Convex optimization problem.

$$E(u; f(w, s_k)) = R(u, h(w_h, s_k)) + \frac{e^{w_\lambda}}{2} \left\| u - g(w_g, s_k) \right\|_2^2 \qquad (1)$$

$$R(u, h(w_h, s_k)) = \min_{v} \alpha_1 \left\| T(h(w_h, s_k))(\nabla_u u - v) \right\|_1 + \alpha_0 \left\| \nabla_v v \right\|_1 . \qquad (2)$$

- Unroll primal-dual optimization scheme [3] on top of deep network.
  - Joint training with deep network.
  - Regularization term parameterized by additional network output.
  - Learning of optimization hyper-parameters for each step.

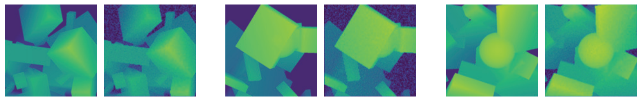
$$\begin{cases} p^{n+1} &= \mathrm{proj}(p^n + \sigma_p \alpha_1 (T(h(w_h, s_k))(\nabla_u \bar{u}^n - \bar{v}^n))) \\ q^{n+1} &= \mathrm{proj}(q^n + \sigma_q \alpha_0 \nabla_v \bar{v}^n) \\ u^{n+1} &= \frac{u^n + \tau_u (\alpha_1 \nabla_u^T T(h(w_h, s_k)) p^{n+1} + e^{w_\lambda} g(w_g, s_k))}{1 + \tau_u e^{w_\lambda}} \\ v^{n+1} &= v^n + \tau_v (\alpha_0 \nabla_v^T q^{n+1} + \alpha_1 T(h(w_h, s_k)) p^{n+1}) \\ \bar{u}^{n+1} &= u^{n+1} + \theta(u^{n+1} - u^n) \\ \bar{v}^{n+1} &= v^{n+1} + \theta(v^{n+1} - v^n) . \end{cases} \qquad (3)$$



Parametrization

Training Data

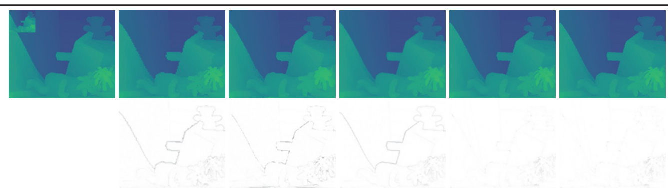
- Generating training data by ray tracing cubes and spheres in 3D.



Evaluation on Noise-Free Data

- Evaluation on Middlebury disparity maps as proposed by [1].

	×2				×4			
	Cones	Teddy	Tsukuba	Venus	Cones	Teddy	Tsukuba	Venus
NN	4.3772	3.2596	9.7968	2.1408	6.1236	4.5168	13.3248	2.9432
Bicubic	3.8392	2.7668	8.3648	1.8192	4.9544	3.5744	10.6960	2.3504
Diebel & Thrun [5]	2.9588	2.1060	6.4208	1.3624	4.5624	3.2040	8.7840	1.9408
Ferstl *et al.* [7]	2.8240	2.1408	7.0592	1.2840	3.6372	2.5068	10.0128	1.4624
Zeyde *et al.* [19]	2.7680	1.9616	6.1936	1.3200	3.8468	2.7812	8.7632	1.7592
Timofte *et al.* [17]	2.7872	1.9816	6.1280	1.3328	3.0256	3.0256	9.6304	1.9616
Aodha *et al.* [1]	4.5076	3.2988	9.6192	2.2088	6.0168	4.1036	13.3328	2.6920
Hornáček *et al.* [10]	3.9744	3.1640	9.2832	2.0592	5.5944	4.7828	11.6352	3.6008
Ferstl *et al.* [8]	2.4988	1.7588	5.6064	1.1464	3.7336	2.6680	7.8416	1.8096
CNN only	1.0275	*0.8201*	*2.3610*	*0.2266*	3.0015	1.5330	*6.4361*	0.4219
CNN + ATGV-L2	*1.0145*	0.8374	**2.3197**	0.2720	*2.9832*	*1.5175*	**6.4223**	*0.4124*
ATGV-Net	**1.0021**	**0.8155**	2.3846	**0.1991**	**2.9293**	**1.5029**	6.6327	**0.3764**

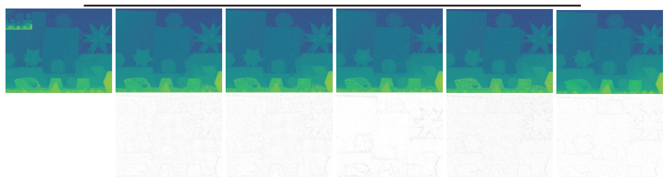


Input & GT   Bicubic   Hornáček *et al.* [10]   Ferstl *et al.* [8]   CNN only   ATGV-Net

Evaluation on Noisy Data

- Middlebury disparity maps with synthetic noise as proposed by [13].

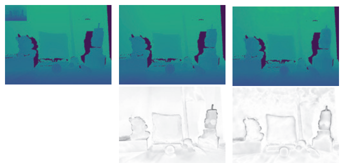
	×2			×4		
	Art	Books	Moebius	Art	Books	Moebius
NN	6.55	6.16	6.59	7.48	6.31	6.78
Bilinear	4.58	3.95	4.20	5.62	4.31	4.56
Yang *et al.* [18]	3.01	1.87	1.92	4.02	2.38	2.42
He *et al.* [9]	3.55	2.37	2.48	4.41	2.74	2.83
Diebel & Thrun [5]	3.49	2.06	2.13	4.51	3.00	3.11
Chan *et al.* [4]	3.44	2.09	2.08	4.46	2.77	2.76
Park *et al.* [13]	3.76	1.95	1.96	4.56	2.61	2.51
Ferstl *et al.* [7]	3.19	1.52	1.47	4.06	*2.21*	*2.03*
CNN only	2.02	1.27	1.50	3.55	2.41	2.68
CNN + ATGV-L2	*1.93*	*1.14*	*1.37*	*3.40*	2.24	2.51
ATGV-Net	**1.84**	**1.13**	**1.24**	**2.98**	**1.72**	**1.95**



Input & GT   He *et al.* [9]   Yang *et al.* [18]   Ferstl *et al.* [7]   CNN only   ATGV-Net

- ToFMark dataset [7].

	Books	Devil	Shark
NN	30.46	27.53	38.21
Bilinear	29.11	25.34	36.34
Kopf *et al.* [12]	27.82	24.30	34.79
He *et al.* [9]	27.11	23.45	33.26
Ferstl *et al.* [7]	**24.00**	*23.19*	*29.89*
ATGV-Net	*24.67*	**21.74**	**28.51**



Input & GT   Ferstl *et al.* [7]   ATGV-Net

References and Acknowledgment

[1] O. M. Aodha, N. D. Campbell, A. Nair, and G. J. Brostow. Patch Based Synthesis for Single Depth Image Super-Resolution. In *ECCV*, 2012.

[2] K. Bredies, K. Kunisch, and T. Pock. Total Generalized Variation. *SIAM Journal on Imaging Sciences*, 3(3):492–526, 2010.

[3] A. Chambolle and T. Pock. A First-Order Primal-Dual Algorithm for Convex Problems with Applications to Imaging. *Journal of Mathematical Imaging and Vision*, 40(1):120–145, 2011.

[4] D. Chan, H. Buisman, C. Theobalt, and S. Thrun. A Noise-aware Filter for Real-time Depth Upsampling. In *ECCV Workshops*, 2008.

[5] J. Diebel and S. Thrun. An Application of Markov Random Fields to Range Sensing. In *NIPS*, 2005.

[6] C. Dong, C. C. Loy, K. He, and X. Tang. Learning a Deep Convolutional Network for Image Super-Resolution. In *ECCV*, 2014.

[7] D. Ferstl, C. Reinbacher, R. Ranftl, M. Rüther, and H. Bischof. Image Guided Depth Upsampling using Anisotropic Total Generalized Variation. In *ICCV*, 2013.

[8] D. Ferstl, M. Rüther, and H. Bischof. Variational Depth Superresolution using Example-Based Edge Representations. In *ICCV*, 2015.

[9] K. He, J. Sun, and X. Tang. Guided Image Filtering. In *ECCV*, 2010.

[10] M. Hornáček, C. Rhemann, M. Gelautz, and C. Rother. Depth Super Resolution by Rigid Body Self-Similarity in 3D. In *CVPR*, 2013.

[11] J. Kim, J. K. Lee, and K. M. Lee. Accurate Image Super-Resolution Using Very Deep Convolutional Networks. In *CVPR*, 2016.

[12] J. Kopf, M. F. Cohen, D. Lischinski, and M. Uyttendaele. Joint Bilateral Upsampling. *TOG*, 26(3):96, 2007.

[13] J. Park, H. Kim, Y.-W. Tai, M. S. Brown, and I.-S. Kweon. High Quality Depth Map Upsampling for 3D-TOF Cameras. In *ICCV*, 2011.

[14] R. Ranftl and T. Pock. A Deep Variational Model for Image Segmentation. In *GCPR*, 2014.

[15] G. Riegler, R. Ranftl, M. Rüther, and H. Bischof. Joint Training of an Convolutional Neural Net and a Global Regression Model. In *BMVC*, 2015.

[16] S. Schulter, C. Leistner, and H. Bischof. Fast and Accurate Image Upscaling with Super-Resolution Forests. In *CVPR*, 2015.

[17] R. Timofte, V. D. Smet, and L. V. Gool. Anchored Neighborhood Regression for Fast Example-Based Super-Resolution. In *ICCV*, 2013.

[18] Q. Yang, R. Yang, J. Davis, and D. Nistér. Spatial-Depth Super Resolution for Range Images. In *CVPR*, 2007.

[19] R. Zeyde, M. Elad, and M. Protter. On Single Image Scale-Up Using Sparse-Representations. In *Curves and Surfaces*, 2010.

This work was supported by *Infineon Technologies Austria AG* and the Austrian Research Promotion Agency under the *FIT-IT Bridge* program, project #838513 (TOFUSION).