Depth Restoration via Joint Training of a Global Regression Model and CNNs

Gernot Riegler*[1]
riegler@icg.tugraz.at
René Ranftl*[1]
ranftl@icg.tugraz.at
Matthias Rüther[1]
ruether@icg.tugraz.at
Thomas Pock[1,2]
pock@icg.tugraz.at
Horst Bischof[1]
bischof@icg.tugraz.at

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

[2] Digital Safety & Security Department
Austrian Institute of Technology
Vienna, Austria

Depth sensors have become increasingly popular in the recent years for a wide range of applications. Among these are video games, gesture control, and applications in the automotive industry. However, the noise and the resolution of depth sensors are problematic. To generate useful depth estimates, the data provided by the sensor is typically subject to post-processing steps. In this work we propose a method that combines a global regression model with a convolutional neural network (CNN) to tackle those problems. Global models, such as TGV-L2 [1, 3], are well suited for the restoration of depth maps, since their prior assumption modeled in the regularization term fit the piecewise affine nature of the data. The data term in these models is designed based on fixed assumptions about the underlying sensor noise, however. Instead of fixed a-priori assumption about the data, we propose to parametrize a global model using a CNN and learn the parameters of the complete system end-to-end in a data-driven way. In this way, our model can automatically adapt to the underlying noise characteristics of the data.

To estimate accurate depth from an input I_k, we use a Global Regression Model (GRM) of the following general form

$$E(u; f(w, I_k)) = R(u, h(w_h, I_k)) + \frac{\exp(w_\lambda)}{2} \|u - g(w_g, I_k))\|^2. \quad (1)$$

Here, $R(u, h(w_h, I_k))$ is a regularization term that introduces prior knowledge and is parameterized by the function $h(w_h, I_k)$. Similarly, the function $g(w_g, I_k)$ is used to transform the input data such that the GRM can make reliable estimates. We use for both parameterization functions h, g a CNN. As regularizer we utilize a non-local pairwise model

$$R(u) = \sum_{i=1}^{N} \sum_{j>i}^{N} n_\varepsilon(h_{ij}(w_h, I_k) \cdot (u_i - u_j)), \quad (2)$$

where n_ε is a twice-differentiable approximation of the Huber function [4]

$$n_\varepsilon(t) = [|t| \leq \varepsilon] \cdot \left(-\tfrac{1}{8\varepsilon^3} t^4 + \tfrac{3}{4\varepsilon} t^2 + \tfrac{3\varepsilon}{8} \right) + [|t| > \varepsilon] \cdot |t| \,. \quad (3)$$

To train the weights $w = [w_h, w_\lambda, w_g]^T$, we assume that K pairs of sensor images I_k together with their ground-truth v_k are given. We formulate the training task as a bi-level problem [6]:

$$\min_{w \in W} \frac{1}{K} \sum_{k=1}^{K} L(u^*(f(w, I_k)), v_k) \quad \text{(HL)}$$

$$\text{s.t. } u^*(f(w, I_k)) = \arg\min_{u \in \mathbb{R}^N} E(u; f(w, I_k)). \quad \text{(LL)}$$

The training procedure can be interpreted as follows: Find parameters w such that the minimizer u^* of $E(\cdot)$ yields a small training loss $L(\cdot)$. We proof in this work necessary conditions for the energy $E(\cdot)$ and the loss $L(\cdot)$ that allows us to efficiently compute the gradient of the higher-level (HL) problem with respect the weights w. If this conditions are satisfied, we can compute the gradient as follows:

$$\Delta E = - \left(\left[\left(\nabla_u^2 E \right)^{-1} \frac{\partial L}{\partial u_k} \right]^T \frac{\partial^2 E}{\partial u_k \partial f} \right) \Bigg|_{u_k = u_k^*}. \quad (4)$$

This formulation fits nicely into the backpropagating scheme to train neural networks and enables us to train the GRM and the CNN in an end-to-end fashion. Further, we can use stochastic gradient descent, or any variation of it, to train the complete model.

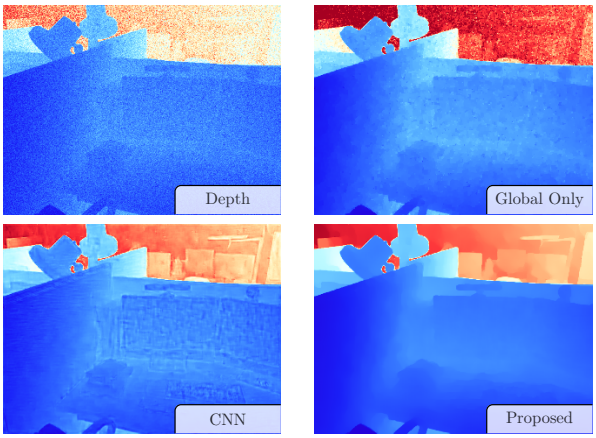

Figure 1: Our proposed method that combines a GRM with a CNN achieves better results than the individual components alone.

		Ours (NL)	CNN	K-SVD	BM3D	TGV-L2
Local Variance	$\sigma_d = 0.5d$	**2.730**	3.913	5.741	3.178	3.042
Poisson	$\lambda = 10^{-3}$	**3.4016**	8.660	6.695	4.129	10.595
Salt & Pepper	$p = 0.35$	**10.484**	18.880	81.685	77.010	80.764

Table 1: Quantitative results on different noise distributions. The error is measured as RMSE in *cm*.

	Ours (NL)	Ours (L)	CNN	Ferstl *et al.*
×2	**2.940**	3.042	6.427	3.834
×4	**4.530**	4.813	8.411	5.506

Table 2: Upscaling results of our method with non-local (NL) and local (L) regularization compared to the SRCNN [2] and the method by Ferstl *et al.* [3]. The error is measured as RMSE in *cm*.

We evaluated our approach for depth map denoising and upscaling on the New Tsukuba dataset [5] for different noise types and levels. An excerp of our quantitative results are presented in Table 1 and 2. Figure 1 depicts an exemplar qualitative result for the denoising task.

The evaluations demonstrate that the proposed GRM parameterized with a CNN is indeed able to adapt to different noise characteristics and performs better than the individual parts independently. Finally, we see potential of our method in computer vision tasks that can benefit from joint learning of a global model and its parameterization.

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

[1] Kristian Bredies, Karl Kunisch, and Thomas Pock. Total Generalized Variation. *SIAM Journal on Imaging Sciences*, 3(3):492–526, 2010.

[2] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Learning a Deep Convolutional Network for Image Super-Resolution. In *ECCV*, 2014.

[3] David Ferstl, Christian Reinbacher, René Ranftl, Matthias Rüther, and Horst Bischof. Image Guided Depth Upsampling using Anisotropic Total Generalized Variation. In *ICCV*, 2013.

[4] Karl Kunisch and Thomas Pock. A Bilevel Optimization Approach for Parameter Learning in Variational Models. *SIAM Journal on Imaging Sciences*, 6(2):938–983, 2013.

[5] Sarah Martull, Martin Peris, and Kazuhiro Fukui. Realistic CG Stereo Image Dataset with Ground Truth Disparity Maps. In *ICPR Workshops*, 2012.

[6] René Ranftl and Thomas Pock. A Deep Variational Model for Image Segmentation. In *GCPR*, 2014.

*Authors contributed equally