Guided Image Super-Resolution: A New Technique for Photogeometric Super-Resolution in Hybrid 3-D Range Imaging

Florin C. Ghesu[1], Thomas Köhler[1,2], Sven Haase[1], and Joachim Hornegger[1,2]

[1] Pattern Recognition Lab, Friedrich-Alexander-Universität Erlangen-Nürnberg
[2] Erlangen Graduate School in Advanced Optical Technologies (SAOT)
{florin.c.ghesu, thomas.koehler}@fau.de

Abstract. In this paper, we augment multi-frame super-resolution with the concept of guided filtering for simultaneous upsampling of 3-D range data and complementary photometric information in hybrid range imaging. Our guided super-resolution algorithm is formulated as joint maximum a-posteriori estimation to reconstruct high-resolution range and photometric data. In order to exploit local correlations between both modalities, guided filtering is employed for regularization of the proposed joint energy function. For fast and robust image reconstruction, we employ iteratively re-weighted least square minimization embedded into a cyclic coordinate descent scheme. The proposed method was evaluated on synthetic datasets and real range data acquired with Microsoft's Kinect. Our experimental evaluation demonstrates that our approach outperforms state-of-the-art range super-resolution algorithms while it also provides super-resolved photometric data.

1 Introduction

3-D range imaging (RI) based on active sensor technologies such as structured light or Time-of-Flight (ToF) cameras is an emerging field of research. Over the past years, with the development of low-cost devices such as Microsoft's Kinect for the consumer market, RI found its way into various computer vision applications [3,12,22] and most recently also to healthcare [2]. Opposed to passive stereo vision approaches, active RI sensors feature the acquisition of dense range images from dynamic scenes in real-time. In addition to range information, complementary photometric data is often provided by the same device in a hybrid imaging system, e.g. color images in case of the Kinect or amplitude data captured by a ToF camera. However, due to technological or economical restrictions, these sensors suffer from a limited spatial resolution which restricts their use for highly accurate measurements. In particular, this is the case for range sensors that may be distorted by random noise and systematic errors depending on the underlying hardware. In order to overcome a limited sensor resolution, image super-resolution has been proposed [14]. A common principle for resolution enhancement is to fuse multiple low-resolution acquisitions with known subpixel displacements into a high-resolution image [7]. The utilized subpixel motion is estimated using image registration methods. For photometric information, the goal of super-resolution is to recover fine structures such as texture barely visible in low-resolution images. In terms of RI, this concept enables accurate 3-D shape scanning [5] which is hard to obtain based on low-resolution range data.

1.1 Related Work

Over the past years, super-resolution has been proposed for a variety of imaging modalities and applications. Traditionally, super-resolution algorithms have been applied to single- or multichannel images encoding photometric information. In many approaches, such as the algorithms proposed by Elad and Feuer [6] or Schultz and Stevenson [18], super-resolution is formulated as maximum a-posteriori (MAP) estimation. In order to consider the presence of outliers, Farsiu et al. [8] introduced a robust extension based on L_1 norm minimization. Babacan et al. [1] have formulated a variational Bayesian approach to estimate high-resolution images and the uncertainty of the underlying model parameters.

In terms of range super-resolution, most prior work adopted techniques originally designed for intensity images. A MAP approach for range super-resolution has been proposed by Schuon et al. [19,20]. Bhavsar and Rajagopalan [4] extended this method by an inpainting scheme to interpolate missing or invalid regions in range data. A Markov Random Field based formulation has been presented by Rajagopalan et al. [17]. Other approaches also exploit complementary photometric data as guidance to reconstruct high-resolution range images. Park et al. [16] utilized adaptive regularization gained from color images to super-resolve range data. A similar technique based on weighted optimization, driven by color images, has been introduced by Schwarz et al. [21]. In the multi-sensor approach proposed by Köhler et al. [10,11], photometric data is utilized as guidance for motion estimation and outlier detection in order to reconstruct reliable high-resolution range images. However, the existence of reliable photometric guidance data required for these methods is not always guaranteed, especially in case of low-cost systems. Furthermore, super-resolution is only applied to a single modality whereas guidance images are not super-resolved.

1.2 Contribution

Opposed to prior work, we propose a novel *guided super-resolution* approach to super-resolve images of two modalities simultaneously. In the context of hybrid RI, we apply our method for photogeometric super-resolution, to reconstruct both, high-resolution range and photometric data. Our algorithm is formulated as joint energy minimization in a MAP framework. In order to exploit correlations between two modalities, we introduce a novel regularizer based on the concept of guided filtering. We employ iteratively re-weighted least square minimization embedded into a cyclic coordinate descent scheme for fast and robust image reconstruction. Our approach is quantitatively and qualitatively evaluated on synthetic data as well as real data captured with Microsoft's Kinect.

2 Photogeometric Super-Resolution Model

Let $\boldsymbol{y}^{(1)}, \ldots, \boldsymbol{y}^{(K)}$ be a sequence of low-resolution input frames, where each frame $\boldsymbol{y}^{(k)}$ is represented by a N_y-dimensional vector. For each input frame $\boldsymbol{y}^{(k)}$, there exists a complementary guidance image $\boldsymbol{p}^{(k)}$ registered to $\boldsymbol{y}^{(k)}$ and denoted as N_p-dimensional vector. In terms of hybrid RI addressed in our work, we use range images as input and corresponding photometric data as guidance. The pair of unknown high-resolution images $(\boldsymbol{x}, \boldsymbol{q})$ with $\boldsymbol{x} \in \mathbb{R}^{N_x}$, $\boldsymbol{q} \in \mathbb{R}^{N_q}$ is related to the low-resolution frames $(\boldsymbol{y}^{(k)}, \boldsymbol{p}^{(k)})$ by a generative model according to:

$$\begin{pmatrix} \boldsymbol{y}^{(k)} \\ \boldsymbol{p}^{(k)} \end{pmatrix} = \begin{pmatrix} \gamma_m^{(k)} \boldsymbol{W}_y^{(k)} & \boldsymbol{0} \\ \boldsymbol{0} & \eta_m^{(k)} \boldsymbol{W}_p^{(k)} \end{pmatrix} \begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{q} \end{pmatrix} + \begin{pmatrix} \gamma_a^{(k)} \boldsymbol{1} \\ \eta_a^{(k)} \boldsymbol{1} \end{pmatrix}, \tag{1}$$

where the system matrices $\boldsymbol{W}_y^{(k)}$, $\boldsymbol{W}_p^{(k)}$ model geometric displacements between $(\boldsymbol{x}, \boldsymbol{q})$ and $(\boldsymbol{y}^{(k)}, \boldsymbol{p}^{(k)})$, as well as the blur induced by the camera point spread function (PSF) and subsampling with respect to the high-resolution image. The model parameters (γ_m, γ_a) and (η_m, η_a) are used to model out-of-plane motion for range data and photometric differences between different guidance images, respectively [10]. Without loss of generality, we model the PSF as a space invariant Gaussian function, to obtain the matrix elements by:

$$W_{m,n} = \exp\left(-\frac{\|\boldsymbol{v}_n - \boldsymbol{u}_m\|_2^2}{2\sigma^2}\right), \tag{2}$$

where $\boldsymbol{v}_n \in \mathbb{R}^2$ are the coordinates of the n^{th} pixel in the high-resolution image, $\boldsymbol{u}_m \in \mathbb{R}^2$ are the coordinates of the m^{th} pixel in the low-resolution frame mapped to the high-resolution grid and σ denotes the width of the PSF.

In order to reconstruct $(\boldsymbol{x}, \boldsymbol{q})$, we propose a joint energy minimization based on a MAP formulation. The objective function consists of a data fidelity term and two regularization terms ensuring the smoothness of the estimates and exploiting correlations between input and guidance images:

$$(\hat{\boldsymbol{x}}, \hat{\boldsymbol{q}}) = \arg\min_{\boldsymbol{x}, \boldsymbol{q}} \left\{ F_{\text{data}}(\boldsymbol{x}, \boldsymbol{q}) + R_{\text{smooth}}(\boldsymbol{x}, \boldsymbol{q}) + R_{\text{correlate}}(\boldsymbol{x}, \boldsymbol{q}) \right\}. \tag{3}$$

2.1 Data Fidelity Term

The data fidelity term measures the similarity between the back-projected high-resolution images $(\boldsymbol{x}, \boldsymbol{q})$ and all low-resolution frames $(\boldsymbol{y}^{(k)}, \boldsymbol{p}^{(k)})$, $k = 1, \ldots, K$, based on our forward model. In order to account for outliers in low-resolution data, we use a weighted L_2 norm error model [11]:

$$F_{\text{data}}(\boldsymbol{x}, \boldsymbol{q}) = \sum_{i=1}^{KN_y} \beta_{y,i} r_{y,i}(\boldsymbol{x})^2 + \sum_{i=1}^{KN_p} \beta_{p,i} r_{p,i}(\boldsymbol{q})^2, \tag{4}$$

where $r_y : \mathbb{R}^{N_x} \to \mathbb{R}^{KN_y}$ and $r_p : \mathbb{R}^{N_q} \to \mathbb{R}^{KN_p}$ denote the residual terms, while $\beta_y \in \mathbb{R}^{KN_y}$ and $\beta_p \in \mathbb{R}^{KN_p}$ represent confidence maps to weight the residuals element-wise. We concatenate the residual terms for all frames according to $r_y(\boldsymbol{x}) = (\boldsymbol{r}_y^{(1)}, \dots, \boldsymbol{r}_y^{(K)})^\top$ and $r_p(\boldsymbol{x}) = (\boldsymbol{r}_p^{(1)}, \dots, \boldsymbol{r}_p^{(K)})^\top$. The residuals of the k^{th} frames are given as:

$$\begin{aligned} \boldsymbol{r}_y^{(k)} &= \boldsymbol{y}^{(k)} - \gamma_m^{(k)} \boldsymbol{W}_y^{(k)} \boldsymbol{x} - \gamma_a^{(k)} \mathbf{1} \\ \boldsymbol{r}_p^{(k)} &= \boldsymbol{p}^{(k)} - \eta_m^{(k)} \boldsymbol{W}_p^{(k)} \boldsymbol{q} - \eta_a^{(k)} \mathbf{1}. \end{aligned} \tag{5}$$

In order to set up our model, we employ a variational approach for optical flow estimation [13] to estimate subpixel motion. Following the multi-sensor approach proposed in [10], motion is estimated on photometric data used as guidance. For range images, $\gamma_m^{(k)}$ and $\gamma_a^{(k)}$ are determined using a range correction scheme [10]. In terms of photometric data, $\eta_m^{(k)}$ and $\eta_a^{(k)}$ are obtained by photometric registration, where $\eta_m^{(k)} = 1$ and $\eta_a^{(k)} = 0$ is set to neglect photometric differences.

2.2 Smoothness Regularization

The smoothness regularization is defined as a sum of regularization terms for the high-resolution images weighted by $\lambda_x, \lambda_q \in \mathbb{R}^+$:

$$R_{\text{smooth}}(\boldsymbol{x}, \boldsymbol{q}) = \lambda_x R(\boldsymbol{x}) + \lambda_q R(\boldsymbol{q}). \tag{6}$$

For regularization, we employ the edge-preserving bilateral total variation (BTV) model [8], defined as:

$$R(\boldsymbol{z}) = \sum_{i=-P}^{P} \sum_{j=-P}^{P} \alpha^{|i|+|j|} \|\boldsymbol{z} - \boldsymbol{S}_v^i \boldsymbol{S}_h^j \boldsymbol{z}\|_1, \tag{7}$$

where $\boldsymbol{S}_v^i$ and $\boldsymbol{S}_h^j$ denote a shift of the image $\boldsymbol{z}$ by i pixels in vertical and j pixels in horizontal direction, and P is a local window size. The shift operators act as derivatives across multiple scales, where α $(0 < \alpha \leq 1)$ is used to control the spatial weighting within the window.

2.3 Interdependence Regularization

We propose a novel interdependence regularization term to exploit local correlations between two modalities. In order to include this kind of prior knowledge, we use a linear, pixel-wise correlation model [9] for the high-resolution images $(\boldsymbol{x}, \boldsymbol{q})$. Assuming that $\boldsymbol{x} \in \mathbb{R}^{N_x}$ and $\boldsymbol{q} \in \mathbb{R}^{N_q}$ have the same dimension $N = N_x = N_q$, interdependence regularization is defined as:

$$R_{\text{correlate}}(\boldsymbol{x}, \boldsymbol{q}) = \lambda_c \|\boldsymbol{x} - \boldsymbol{A}\boldsymbol{q} - \boldsymbol{b}\|_2^2, \tag{8}$$

where $\lambda_c \in \mathbb{R}^+$ weights the correlation between $\boldsymbol{x}$ and $\boldsymbol{q}$. The filter coefficients $\boldsymbol{A} \in \mathbb{R}^{N \times N}$ and $\boldsymbol{b} \in \mathbb{R}^N$ are constructed to model this correlation. The higher λ_c is chosen, the higher the correlation between input and guidance images is enforced. In case of $N_x \neq N_q$, we use bicubic interpolation to reshape $\boldsymbol{x}$ and $\boldsymbol{q}$ to the same size in order to compute $R_{\text{correlate}}(\boldsymbol{x}, \boldsymbol{q})$.

3 Numerical Optimization Algorithm

A direct minimization of the joint energy function in Eq. (3) requires a simultaneous estimation of $\boldsymbol{x}$ and $\boldsymbol{q}$ with the associated confidence maps $\boldsymbol{\beta}_y$ and $\boldsymbol{\beta}_p$ as well as of the filter coefficients $\boldsymbol{A}$ and $\boldsymbol{b}$. However, this is a highly ill-posed and computationally demanding inverse problem. For a fast and robust solution, we utilize an iterative re-weighted least square (IRLS) minimization [11] embedded into a cyclic coordinate descent scheme. We decompose the estimation of $\boldsymbol{x}$ and $\boldsymbol{q}$ into a sequence of n least square optimization problems to determine $\boldsymbol{x}^{(t)}$ and $\boldsymbol{q}^{(t)}$ at iteration $t = 1, \ldots, n$. Simultaneously to image reconstruction, the confidence maps $\boldsymbol{\beta}_y^{(t)}$ and $\boldsymbol{\beta}_p^{(t)}$ as well as the filter coefficients $\boldsymbol{A}^{(t)}$ and $\boldsymbol{b}^{(t)}$ are determined analytically and refined at each iteration. In detail, our optimization is performed as follows.

Confidence Map Computation. Let $(\boldsymbol{x}^{(t)}, \boldsymbol{q}^{(t)})$ be the estimates for the high-resolution images at iteration t and $(\boldsymbol{r}_y^{(t)}, \boldsymbol{r}_p^{(t)})$ be the associated residual error computed according to Eq. (5). Then, following [11] the confidence maps $\beta_y^{(t)}$ and $\beta_p^{(t)}$ at iteration t are derived analytically according to:

$$\beta_{y,i}^{(t)} = \begin{cases} 1 & \text{if } |\boldsymbol{r}_{y,i}^{(t)}| \leq \epsilon_y \\ \frac{\epsilon_y}{|\boldsymbol{r}_{y,i}^{(t)}|} & \text{otherwise} \end{cases} \quad \beta_{p,i}^{(t)} = \begin{cases} 1 & \text{if } |\boldsymbol{r}_{p,i}^{(t)}| \leq \epsilon_p \\ \frac{\epsilon_p}{|\boldsymbol{r}_{p,i}^{(t)}|} & \text{otherwise} \end{cases}, \tag{9}$$

where ϵ_y and ϵ_p are initialized by the standard deviations of $\boldsymbol{r}_y^{(t)}$ and $\boldsymbol{r}_p^{(t)}$, respectively. For outlier detection, this scheme assigns a smaller confidence to low-resolution observations that result in higher residual errors.

Guidance Image Super-Resolution. In order to estimate $\boldsymbol{q}^{(t)}$, we solve Eq. (3) with respect to $\boldsymbol{q}$ using the confidence map $\boldsymbol{\beta}_p^{(t)}$ while keeping $\boldsymbol{x}$ fixed. For this step, we employ interdependence regularization in a non-symmetric way, as super-resolution of guidance data would not benefit from the complementary input images. The updated estimate $\boldsymbol{q}^{(t)}$ is obtained according to:

$$\boldsymbol{q}^{(t)} = \arg\min_{\boldsymbol{q}} \left\{ F_{\text{data}}(\boldsymbol{x}, \boldsymbol{q}) + R_{\text{smooth}}(\boldsymbol{x}, \boldsymbol{q}) \right\}_{\boldsymbol{x}=\boldsymbol{x}^{(t-1)}} . \tag{10}$$

For the solution of this convex optimization problem, scaled conjugate gradient (SCG) iterations [15] are used to determine $\boldsymbol{q}^{(t)}$ using $\boldsymbol{q}^{(t-1)}$ as initial guess.

Guided Filtering. Once the guidance image $\boldsymbol{q}^{(t)}$ and the super-resolved image $\boldsymbol{x}^{(t-1)}$ from the previous iteration are determined, the filter coefficients $\boldsymbol{A}^{(t)}$ and $\boldsymbol{b}^{(t)}$ are updated. In order to exploit the correlation between $\boldsymbol{q}^{(t)}$ and $\boldsymbol{x}^{(t-1)}$, we keep both fixed and determine the filter coefficients analytically using guided

Algorithm 1 Guided image super-resolution algorithm

1: **for** $t = 1 \dots t_{max}$ **do** ▷ t_{max}: maximum number of iterations
2: Update confidence maps $\beta_y^{(t)}$ and $\beta_p^{(t)}$ according to Eq. (9).
3: Estimate high-resolution photometric data $q^{(t)}$ at step t according to Eq. (10):

$$q^{(t)} = \arg\min_{\boldsymbol{q}} \sum_{i=1}^{KN_p} \beta_{p,i}^{(t)} r_{p,i}^{(t)}(\boldsymbol{q})^2 + \lambda_q R(\boldsymbol{q})$$

4: Determine $\boldsymbol{A}^{(t)}$ and $\boldsymbol{b}^{(t)}$ from $\boldsymbol{q}^{(t)}$ and $\boldsymbol{x}^{(t-1)}$ according to Eq. (11) and Eq. (12).
5: Estimate high-resolution range image $x^{(t)}$ at step t according to Eq. (13):

$$x^{(t)} = \arg\min_{\boldsymbol{x}} \sum_{i=1}^{KN_y} \beta_{y,i}^{(t)} r_{y,i}^{(t)}(\boldsymbol{x})^2 + \lambda_x R(\boldsymbol{x}) + \lambda_c \|\boldsymbol{x} - \boldsymbol{A}^{(t)}\boldsymbol{q}^{(t)} - \boldsymbol{b}^{(t)}\|_2^2$$

6: If not converged proceed with next iteration

filtering [9]. Omitting the iteration index for sake of clarity, we construct $\boldsymbol{A}$ as diagonal matrix and calculate the filter coefficients according to:

$$\tilde{A}_{k,k} = \frac{\frac{1}{|\omega_k|}\sum_{i\in\omega_k} q_i x_i - \mathrm{E}_{\omega_k}(\boldsymbol{q})\mathrm{E}_{\omega_k}(\boldsymbol{x})}{\mathrm{Var}_{\omega_k}(\boldsymbol{q}) + \epsilon} \tag{11}$$

$$\tilde{b}_k = \mathrm{E}_{\omega_k}(\boldsymbol{x}) - \tilde{A}_{k,k}\mathrm{E}_{\omega_k}(\boldsymbol{q}), \tag{12}$$

where ω_k denotes a local neighborhood of size $|\omega_k|$ and radius r centered at the k^{th} pixel in $(\boldsymbol{x}, \boldsymbol{q})$, ϵ is a regularization factor for guided filtering, and $\mathrm{E}_{\omega_k}(\cdot)$ and $\mathrm{Var}_{\omega_k}(\cdot)$ are the mean and variance in ω_k. The filter coefficients $\boldsymbol{A}$ and $\boldsymbol{b}$ are computed by box filtering of $\tilde{\boldsymbol{A}}$ and $\tilde{\boldsymbol{b}}$ using the window defined by ω_k.

Input Image Super-Resolution. Finally, we solve Eq. (3) with respect to $\boldsymbol{x}$ to obtain a refined estimate $\boldsymbol{x}^{(t)}$ under the guidance of $\boldsymbol{q}^{(t)}$. This is achieved by means of interdependence regularization based on $\boldsymbol{A}^{(t)}$ and $\boldsymbol{b}^{(t)}$ according to:

$$\boldsymbol{x}^{(t)} = \arg\min_{\boldsymbol{x}} \{F_{\mathrm{data}}(\boldsymbol{x}, \boldsymbol{q}) + R_{\mathrm{smooth}}(\boldsymbol{x}, \boldsymbol{q}) + R_{\mathrm{correlate}}(\boldsymbol{x}, \boldsymbol{q})\}_{\boldsymbol{q}=\boldsymbol{q}^{(t)}} . \tag{13}$$

In the same way as for the guidance images, the resulting convex optimization problem is solved employing SCG, where $\boldsymbol{x}^{(t-1)}$ is used as initial guess. The outline of the proposed algorithm is depicted in Table 1.

4 Experiments and Results

We compared the proposed guided super-resolution (GSR) approach to MAP super-resolution using the L_2 norm [6] and L_1 norm model [8] working on a single modality. For photometric data, all methods were directly applied to intensity images using optical flow [13] for motion estimation. For range super-resolution,

Table 1. Peak-signal-to-noise ratio (PSNR) and structural similarity (SSIM, in brackets) for synthetic data averaged over ten sequences. We compared bicubic interpolation (second column), MAP super-resolution using the L_1 norm (third column) as well as the L_2 norm (forth column) to the proposed guided super-resolution (GSR, last column).

	Sequence	Interpolation	MAP - L_1	MAP - L_2	GSR
Range	Bunny-1	32.78 (0.96)	34.10 (0.96)	34.05 (0.97)	**35.01 (0.98)**
Range	Bunny-2	31.29 (0.94)	32.84 (0.95)	33.22 (0.97)	**33.34 (0.98)**
Range	Dragon-1	24.63 (0.57)	27.68 (0.72)	28.71 (0.84)	**30.00 (0.91)**
Range	Dragon-2	27.14 (0.75)	29.09 (0.84)	29.76 (0.93)	**30.80 (0.95)**
Photo.	Bunny-1	28.48 (0.79)	29.82 (0.87)	29.79 (0.87)	**29.79 (0.88)**
Photo.	Bunny-2	30.05 (0.81)	31.35 (0.86)	31.42 (0.86)	**31.43 (0.86)**
Photo.	Dragon-1	23.34 (0.65)	24.25 (0.72)	24.24 (0.71)	**24.27 (0.72)**
Photo.	Dragon-2	24.65 (0.66)	25.60 (0.72)	25.51 (0.70)	**25.60 (0.72)**

we employed the multi-sensor super-resolution approach [10] to derive the motion estimate from the corresponding photometric data. Super-resolution was applied in a sliding window scheme using $K = 31$ successive frames, where the central frame was chosen as reference for motion estimation. We used a magnification factor of $f = 4$ for range and photometric data. The PSF width was approximated to $\sigma_p = 0.4$ for photometric data and $\sigma_y = 0.6$ for range data. For BTV regularization, we set $\alpha = 0.7$ and $P = 1$. For guided filtering, we set $\epsilon = 10^{-4}$ and $r = 1$. The regularization weights were optimized using a grid search on a training data set with known ground truth and were set to $\lambda_x = 0$, $\lambda_q = 0.002$ and $\lambda_c = 1$ for all experiments. We used SCG with a termination tolerance of 10^{-3} for the pixels of $(\boldsymbol{x}, \boldsymbol{q})$ and the objective function value. The maximum number of SCG iterations was set to 50 for 15 IRLS iterations[3].

4.1 Synthetic Hybrid Range Data

We simulated synthetic range data with known ground truth using a RI simulator [23]. The ground truth was generated in VGA resolution (640×480 px) using RGB images to encode photometric information. All low-resolution frames were downsampled by a factor of $f = 4$. Range data was affected by distance-dependent Gaussian noise ($\sigma_n = 8\,\mathrm{mm}$) and Gaussian blur ($\sigma_b = 3\,\mathrm{mm}$). Photometric data was disturbed by space invariant Gaussian noise ($\sigma_n = 5 \cdot 10^{-4}$). We generated random displacements of consecutive frames to simulate camera movements. The quality of super-resolved data with respect to the ground truth was assessed using the peak-signal-to-noise ratio (PSNR) and structural similarity (SSIM). For RGB images, this evaluation was performed on the gray-scale converted color images. Table 1 shows the PSNR and SSIM measures averaged over ten subsequent sequences in sliding window processing. Qualitative results are depicted in Fig. 1 for range images and Fig. 2 for photometric data.

[3] Supplementary material is available at http://www5.cs.fau.de/research/data/

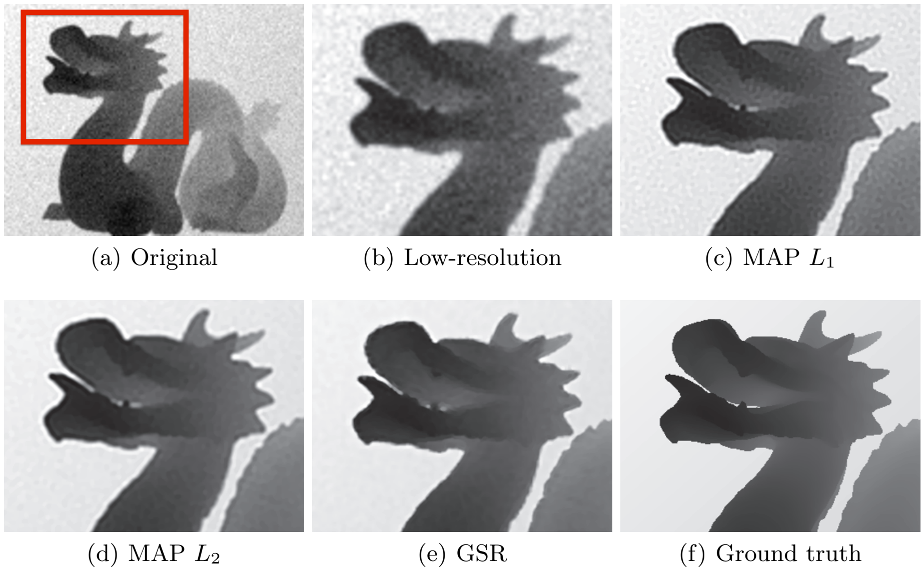


Fig. 1. Results for simulated *Dragon-2* dataset: Low-resolution range data (a) and selected region of interest (b), results of MAP super-resolution using the L_1 (c) and L_2 norm model (d), and our proposed GSR method (e) compared to the ground truth (f).

4.2 Microsoft's Kinect Datasets

We acquired real range data using a Microsoft Kinect device. Range and photometric data was captured in VGA resolution (640×480 px) using a frame rate of 30 fps. Color images in the RGB color space were used to encode photometric data. We applied super-resolution to the gray-scale converted color images for visual comparison. During the acquisition, the device was held in the hand such that a small shaking ensured the required motion for super-resolution over consecutive frames. We used the same parameter settings as for synthetic data. Qualitative results are shown for different acquisitions in Fig. 3 and Fig. 4. Finally, we also rendered 3-D meshes with a texture overlay from super-resolved range data. A comparison of the different approaches is depicted in Fig. 5.

5 Discussion

In this work, we introduce a novel method for photogeometric resolution enhancement based on the concept of guided super-resolution. Unlike prior work, our method super-resolves range and photometric data simultaneously. This allows us to exploit photometric data as guidance for range super-resolution within a joint framework. Our experimental evaluation demonstrates the performance of our method on real as well as synthetic data. In case of range images, we

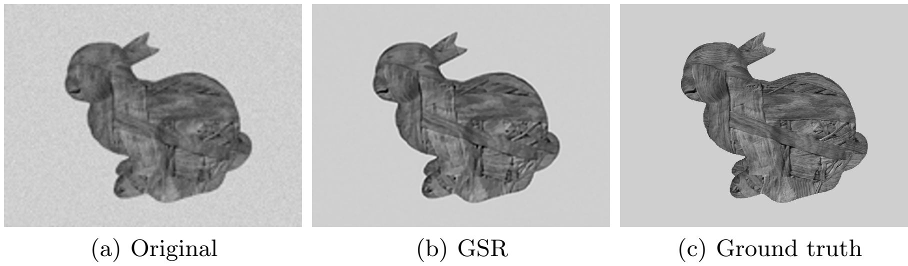


Fig. 2. Results for simulated *Bunny-1* dataset: Low-resolution photometric data (a) and result of guided super-resolution (GSR) (b) compared to the ground truth (c).

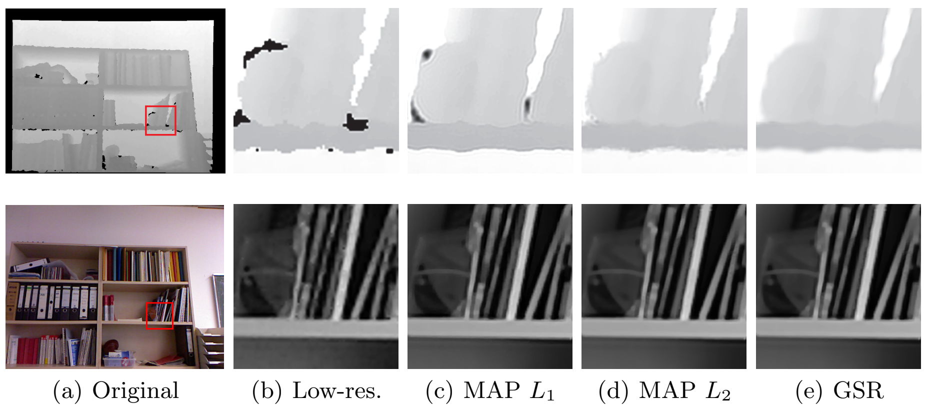
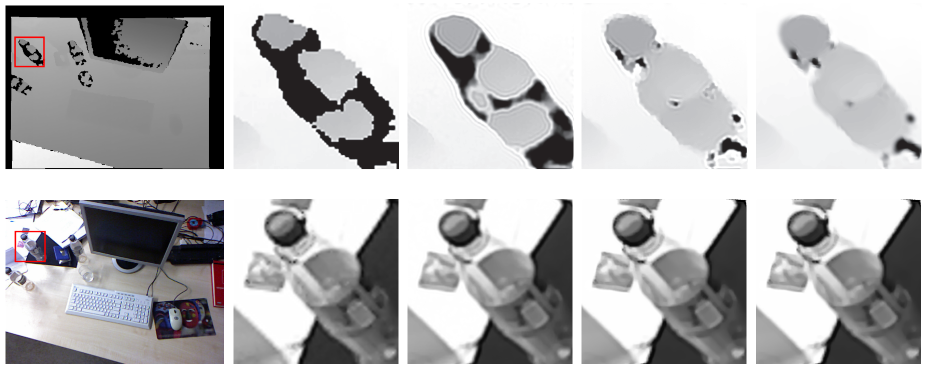


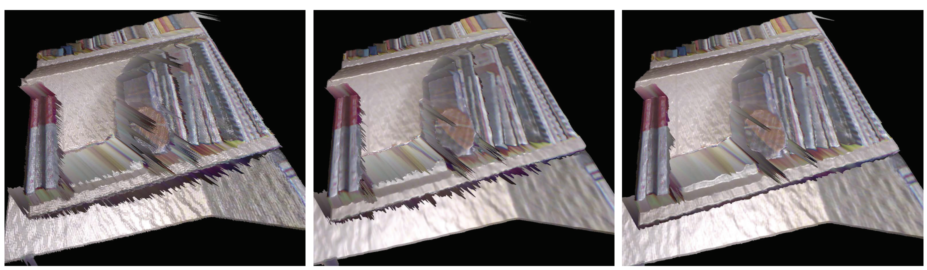
Fig. 3. Real data set example showing range data (first row) and photometric data (second row): Low-resolution range and RGB frame (a), selected low-resolution region of interest (b), results for MAP super-resolution using L_1 norm (c) and L_2 norm (d), and the super-resolved images using the proposed guided super-resolution (GSR) (e).

achieved an improvement of $\sim 1\,\text{dB}$ for PSNR and ~ 0.04 for the SSIM measure compared to super-resolution applied only on range data (see Table 1). On photometric data, our method achieved similar performance as other state-of-the-art algorithms. However, this behavior was expected since the photometric data is not guided by range images in our formulation. Visual inspection using Kinect acquisitions demonstrates the benefits of our approach for real data. For range images, we observed an improved trade-off between edge reconstruction and smoothing in flat regions as depicted in Fig. 3 and Fig. 4. This is caused by the proposed interdependence regularization, which exploits local correlation between modalities. Additionally, invalid range pixels are corrected as complementary information from multiple frames is fused in image reconstruction.



(a) Original (b) Low-res. (c) MAP L_1 (d) MAP L_2 (e) GSR

Fig. 4. Real data set example showing range data (first row) and photometric data (second row): Low-resolution range and RGB frame (a), selected low-resolution region of interest (b), results for MAP super-resolution using L_1 norm (c) and L_2 norm (d), and the super-resolved images using the proposed guided super-resolution (GSR) (e).



(a) MAP L_1 (b) MAP L_2 (c) GSR

Fig. 5. 3-D meshes with texture overlay triangulated from super-resolved range data.

6 Conclusion and Future Work

This paper proposes guided super-resolution to super-resolve images of two complementary modalities within a joint framework. We applied this concept to reconstruct high-resolution range and photometric data in hybrid range imaging. Our method exploits local correlations between both modalities for a novel interdependence regularization based on guided filtering. Experiments on real and synthetic images demonstrate the performance of our approach in comparison to methods working solely on one modality. In our future work, we will investigate the applicability of our method to different RI setups, e. g. for ToF imaging to acquire range and amplitude data. As the proposed interdependence regularization is independent of the utilized modalities, its adoption to different multi-sensor setups such as RGB or multispectral imaging seems attractive.

Acknowledgments. The authors gratefully acknowledge funding of the Erlangen Graduate School in Advanced Optical Technologies (SAOT) by the German National Science Foundation (DFG) in the framework of the excellence initiative and the support by the DFG under Grant No. HO 1791/7-1.

References

1. Babacan, S.D., Molina, R., Katsaggelos, A.K.: Variational Bayesian Super Resolution. IEEE Transactions on Image Processing 20(4), 984–999 (2011)
2. Bauer, S., Seitel, A., Hofmann, H., Blum, T., Wasza, J., Balda, M., Meinzer, H.P., Navab, N., Hornegger, J., Maier-Hein, L.: Real-Time Range Imaging in Health Care: A Survey. In: Time-of-Flight and Depth Imaging. Sensors, Algorithms, and Applications, pp. 228–254. Springer (2013)
3. Beder, C., Bartczak, B., Koch, R.: A Comparison of PMD-Cameras and Stereo-Vision for the Task of Surface Reconstruction using Patchlets. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 1–8 (2007)
4. Bhavsar, A.V., Rajagopalan, A.N.: Range map superresolution-inpainting, and reconstruction from sparse data. Computer Vision and Image Understanding 116(4), 572–591 (2012)
5. Cui, Y., Schuon, S., Chan, D., Thrun, S., Theobalt, C.: 3D shape scanning with a time-of-flight camera. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 1173–1180 (2010)
6. Elad, M., Feuer, A.: Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images. IEEE Transactions on Image Processing 6(12), 1646–1658 (1997)
7. Farsiu, S., Robinson, D., Elad, M., Milanfar, P.: Advances and Challenges in Super-Resolution. International Journal of Imaging Systems and Technology 14, 47–57 (2004)
8. Farsiu, S., Robinson, D., Elad, M., Milanfar, P.: Fast and robust multiframe super resolution. IEEE Transactions on Image Processing 13(10), 1327–1344 (2004)
9. He, K., Sun, J., Tang, X.: Guided Image Filtering. In: Proceedings of the 11th European Conference on Computer Vision: Part I. pp. 1–14. Springer (2010)
10. Köhler, T., Haase, S., Bauer, S., Wasza, J., Kilgus, T., Maier-Hein, L., Feubner, H., Hornegger, J.: ToF Meets RGB: Novel Multi-Sensor Super-Resolution for Hybrid 3-D Endoscopy. Medical Image Computing and Computer Assisted Intervention 16, 139–146 (2013)
11. Köhler, T., Haase, S., Bauer, S., Wasza, J., Kilgus, T., Maier-Hein, L., Feuner, H., Hornegger, J.: Outlier Detection for Multi-Sensor Super-Resolution in Hybrid 3D Endoscopy. In: Bildverarbeitung für die Medizin 2014, pp. 84–89. Springer (2014)
12. Kurmankhojayev, D., Hasler, N., Theobalt, C.: Monocular Pose Capture with a Depth Camera Using a Sums-of-Gaussians Body Model. In: Pattern Recognition, vol. 8142, pp. 415–424. Springer (2013)
13. Liu, C.: Beyond Pixels: Exploring New Representations and Applications for Motion Analysis. Ph.D. thesis, Massachusetts Institute of Technology (2009)
14. Milanfar, P.: Super-resolution imaging. CRC Press (2010)
15. Nabney, I.T.: NETLAB: Algorithms for Pattern Recognition. Advances in Pattern Recognition, Springer, 1^{st} edn. (2002)
16. Park, J., Kim, H., Tai, Y., Brown, M., Kweon, I.: High quality depth map upsampling for 3D-TOF cameras. In: International Conference on Computer Vision. pp. 1623–1630 (2011)

17. Rajagopalan, A., Bhavsar, A., Wallhoff, F., Rigoll, G.: Resolution Enhancement of PMD Range Maps. Proceedings of the 30th DAGM symposium on Pattern Recognition 5096, 304–313 (2008)
18. Schultz, R.R., Stevenson, R.L.: Extraction of high-resolution frames from video sequences. IEEE Transaction on Image Processing 5, 996–1011 (1996)
19. Schuon, S., Theobalt, C., Davis, J., Thrun, S.: High-Quality Scanning Using Time-Of-Flight Depth Superresolution. IEEE Conference on Computer Vision and Pattern Recognition 1, 1–7 (2008)
20. Schuon, S., Theobalt, C., Davis, J., Thrun, S.: LidarBoost: Depth Superresolution For ToF 3D Shape Scanning. IEEE Conference on Computer Vision and Pattern Recognition 1, 343–350 (2009)
21. Schwarz, S., Sjostrom, M., Olsson, R.: A Weighted Optimization Approach to Time-of-Flight Sensor Fusion. IEEE Transactions on Image Processing 23(1), 214–225 (Jan 2014)
22. Shotton, J., Girshick, R., Fitzgibbon, A., Sharp, T., Cook, M., Finocchio, M., Moore, R., Kohli, P., Criminisi, A., Kipman, A., Blake, A.: Efficient Human Pose Estimation from Single Depth Images. Pattern Analysis and Machine Intelligence 35(12), 2821–2840 (2013)
23. Wasza, J., Bauer, S., Haase, S., Schmid, M., Reichert, S., Hornegger, J.: RITK: The Range Imaging Toolkit - A Framework for 3-D Range Image Stream Processing. In: VMV. pp. 57–64. Eurographics Association (2011)