Multi-Sensor Super-Resolution for Hybrid Range Imaging with Application to 3-D Endoscopy and Open Surgery

Thomas Köhler[a,b,*], Sven Haase[a], Sebastian Bauer[a], Jakob Wasza[a], Thomas Kilgus[c], Lena Maier-Hein[c], Christian Stock[d], Joachim Hornegger[a,b], Hubertus Feußner[e]

[a]*Pattern Recognition Lab, Department of Computer Science, Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany*
[b]*Erlangen Graduate School in Advanced Optical Technologies (SAOT), Germany*
[c]*Division of Medical and Biological Informatics Junior Group: Computer-assisted Interventions, German Cancer Research Center (DKFZ) Heidelberg, Germany*
[d]*Institute of Medical Biometry and Informatics, University of Heidelberg, Germany*
[e]*Research Group Minimally-invasive interdisciplinary therapeutical intervention, Klinikum rechts der Isar of the Technical University Munich, Germany*

Abstract

In this paper, we propose a multi-sensor super-resolution framework for hybrid imaging to super-resolve data from one modality by taking advantage of additional guidance images of a complementary modality. This concept is applied to hybrid 3-D range imaging in image-guided surgery, where high-quality photometric data is exploited to enhance range images of low spatial resolution. We formulate super-resolution based on the maximum a-posteriori (MAP) principle and reconstruct high-resolution range data from multiple low-resolution frames and complementary photometric information. Robust motion estimation as required for super-resolution is performed on photometric data to derive displacement fields of subpixel accuracy for the associated range images. For improved reconstruction of depth discontinuities, a novel adaptive regularizer exploiting correlations between both modalities is embedded to MAP estimation. We evaluated our method on synthetic data as well as ex-vivo images in open surgery and endoscopy. The proposed multi-sensor framework improves the peak signal-to-noise ratio by 2 dB and structural similarity by 0.03 on average compared to conventional single-sensor approaches. In ex-vivo experiments on porcine organs, our method achieves substantial improvements in terms of depth discontinuity reconstruction.

Keywords: Hybrid range imaging, Sensor fusion, Super-resolution, 3-D endoscopy, Open surgery

1. Introduction

Hybrid imaging is an emerging field of research in medical imaging describing the fusion of different modalities. From a general perspective, sensor fusion enables an augmented representation of complementary information. Common setups are the combination of positron emission tomography (PET) in nuclear medicine with computed tomography (CT) or magnet resonance imaging (MRI) to visualize metabolism and internal structures simultaneously. A novel hybrid imaging setup addressed in this paper combines range imaging (RI) technologies with RGB sensors to augment 3-D range data with photometric information.

1.1. Applications

In recent years, RI has been proposed for several applications in healthcare (Bauer et al., 2013). In this work, we examine two different applications of hybrid RI in the field of image-guided surgery.

In terms of minimally invasive procedures, various approaches to gain intra-operative range data have been introduced and evaluated with respect to their clinical usability (Maier-Hein et al., 2014). Stereoscopy (Field et al., 2009) has been proposed as a passive technique to capture 3-D range data for interventional imaging. On the other hand, active sensor technologies based on structured light

*Corresponding author. Telephone: +49-9131-85-27775; Fax: +49-9131-303811
Email address: `thomas.koehler@fau.de` (Thomas Köhler)

(SL) (Schmalz et al., 2012) or Time-of-Flight (ToF) (Penne et al., 2009) have been examined. These sensors can be augmented with photometric information to enable hybrid RI within one single endoscope, e. g. to facilitate ToF/RGB endoscopy (Haase et al., 2013a). This fusion of complementary modalities provides the surgeon a comprehensive view of a scene including 2-D and 3-D information. In addition, sensor fusion is beneficial to enhance robustness and reliability of many image processing tasks, e. g. for localization and tracking of laparoscopic instruments (Haase et al., 2013b).

In open surgery, one common workflow is to register pre-operative 3-D planning data acquired, e. g. by CT, with intra-operative range data gained by means of marker-less RI technologies (Bauer et al., 2013). As for minimally invasive procedures, stereo vision is a common technique for intra-operative imaging evaluated, e. g. for brain shift compensation in image-guided neurosurgery (Sun et al., 2005). In the field of active sensor technologies, laser scanning (Cash et al., 2007) and ToF (Mersmann et al., 2011; Kilgus et al., 2014) have been introduced as marker-less approaches to facilitate augmented reality. Similarly to 3-D endoscopy, range data acquired in open surgery can be augmented by photometric information to enhance the intuitive representation of the underlying scene.

1.2. Technical Challenges

As demonstrated in recent studies (Maier-Hein et al., 2014), active and passive approaches are complementary RI technologies. While passive stereo vision is able to provide highly accurate range information under ideal situations, it might be error-prone on surfaces without texture or with repetitive structures. Active sensors are less influenced by texture but are limited in their spatial resolution and the signal-to-noise ratio (SNR) due to inherent physical or economical limitations. In particular, the resolutions of low-cost ToF or SL sensors are rather low compared to photometric information. For minimally invasive procedures or open surgery, this means a major limitation and restricted the integration of RI to many clinical workflows.

In order to enhance the spatial resolution of digital images, *super-resolution* (Milanfar, 2010) is a technique to reconstruct high-resolution (HR) data from the acquired raw images. One common approach is to fuse multiple low-resolution (LR) frames into a new HR image (Greenspan, 2008). Most conventional super-resolution algorithms exploit only images from a single modality and are termed as *single-sensor* methods below. Opposed to this approach, *multi-sensor* techniques take advantage of additional guidance by multiple modalities. In terms of RI, high-quality photometric data can be exploited for range super-resolution.

1.3. Contribution and Outline

We propose a multi-sensor super-resolution framework and present its application for hybrid RI in image-guided surgery. This paper is an extension of a conference proceeding (Köhler et al., 2013) introducing this concept for 3-D endoscopy for the first time. In this work, we examine sensor data fusion of range and photometric information for two RI setups applicable to 3-D endoscopy and open surgery as generalization of (Köhler et al., 2013). This concept is used to derive subpixel displacement fields for range super-resolution under the guidance of photometric information. As extension of our prior work, we also introduce a novel adaptive regularization technique, where photometric data is exploited for improved edge reconstruction in range super-resolution. In our experimental evaluation, we demonstrate the performance of our method on ex-vivo porcine organs qualitatively and quantitatively.

The remainder of this paper is organized as follows: Sect. 2 discusses relevant work on upsampling of range data. Sect. 3 introduces our system calibration approach to perform sensor data fusion for multi-sensor super-resolution. In Sect. 4, we introduce MAP super-resolution as computational framework for our approach. Sect. 5 introduces our super-resolution approach for hybrid RI. In Sect. 6, our method is evaluated for image-guidance in minimally invasive as well as open surgery. A discussion of our method is given in Sect. 7

2. Related Work

Multi-frame super-resolution algorithms exploit relative movements of a camera with respect to a 3-D surface. Due to subpixel displacements in the associated image sequence, an HR image of finer spatial sampling can be reconstructed (Irani and Peleg, 1991). Unlike single-frame upsampling techniques (Kopf et al., 2007; He et al., 2010; Park et al., 2011), image deblurring and denoising may be treated in a joint approach. In this paper, we distinguish between single-sensor and multi-sensor methods.

2.1. Single-Sensor Approach

Early approaches divide super-resolution in a motion estimation and a reconstruction stage formulated in the frequency domain (Tsai and Huang, 1984) or the spatial domain using set theoretic or statistical parameter estimation (Elad and Feuer, 1997; Farsiu et al., 2004). The latter approaches are attractive since prior knowledge can be easily incorporated to regularize the reconstruction process. However, most approaches are based on simplified motion models, e. g. a rigid transformation *in the image plane* ignoring the projective mapping of a real camera. For this reason and to handle non-rigid motion, optical flow estimation has been also widely employed (Zhao and Sawhney, 2002; Mitzel et al., 2009). As accurate motion estimation is a major challenge for super-resolution, improved registration schemes have been proposed (Vandewalle et al., 2006). Similarly, image reconstruction and motion estimation can be treated in a *joint* optimization procedure, e. g. based on joint MAP or Bayesian approaches (Hardie et al., 1997; Pickup et al., 2007), variable projection (Robinson et al., 2009) or expectation maximization (Fransens et al., 2007). Even if the reliability of super-resolution is improved, joint optimization is a computationally demanding and highly non-convex problem (Pickup et al., 2007) or limited to simplified motion models.

Recently, such methods have been adopted to RI by Schuon et al. (2009) or Bhavsar and Rajagopalan (2012). However, these techniques super-resolve a single modality, i. e. range data, without exploiting sensor fusion with guidance images, i. e. photometric data. Our work shows that sensor fusion can be employed to enhance super-resolution reconstruction in terms of motion estimation as well as regularization in statistical, spatial domain algorithms.

2.2. Multi-Sensor Approach

There are only few multi-sensor methods, which are often limited to specific applications. Multi-sensor super-resolution for single RGB/infrared images has been proposed by Zomet and Peleg (2002). One approach designed for PET/CT scanners has been proposed by Kennedy et al. (2007), where PET scans are super-resolved and augmented by anatomical information gained from high-resolution CT scans. Even if this approach is beneficial for PET resolution enhancement, it does not generalize to different applications. In terms of hybrid 3-D endoscopy, a first multi-sensor super-resolution framework has been introduced, recently (Köhler et al., 2013). In this approach, photometric information is employed as guidance for range super-resolution. Motion estimation is performed on high-quality photometric data in order to derive accurate displacement fields for range images. However, this framework is based on a simplified setup for sensor fusion exploiting a homographic mapping and has not been investigated for more general systems, yet.

This work generalizes the approach of Köhler et al. (2013) and demonstrates its application in minimally invasive and open surgery.

3. System Calibration

Our framework exploits sensor data fusion between range images and complementary photometric information. The LR range data of size $M_1 \times M_2$ is denoted as $\boldsymbol{Y}$ defined on a rectangular grid $\Omega_r \subset \mathbb{R}^2$. We consider a sequence of K range images denoted as $\boldsymbol{Y}^{(1)}, \ldots, \boldsymbol{Y}^{(K)}$. For convenience, each image $\boldsymbol{Y}$ is linearized to a vector $\boldsymbol{y} \in \mathbb{R}^M$ with $M = M_1 \cdot M_2$. For each $\boldsymbol{Y}^{(k)}$ there exist an associated image $\boldsymbol{z}^{(k)} \equiv \boldsymbol{Z}^{(k)}$ of size $L_1 \times L_2$ defined on domain $\Omega_z \subset \mathbb{R}^2$ encoding photometric information. Neglecting occlusions, each pixel $\boldsymbol{u}_z \in \Omega_z$ is related to one pixel $\boldsymbol{u}_r \in \Omega_r$ in the corresponding range image according to a mapping $f : \Omega_z \to \Omega_r$. We consider only static setups with fixed relative positions between both sensors and thus f is assumed to be fixed. Our aim is to estimate f and align each $\boldsymbol{z}^{(k)}$ to $\boldsymbol{y}^{(k)}$ such that a pair $(\boldsymbol{u}_z, \boldsymbol{u}_r)$ of corresponding pixels relates to the same 3-D scene point $\boldsymbol{U} \in \mathbb{R}^3$ (see Fig. 1). Please note that f is defined up to a scale factor to preserve the resolution of photometric data with respect to range images.

We examine the system calibration of two hybrid RI setups to perform sensor fusion.

3.1. Homographic Mapping

In the first setup, a 3-D surface is acquired by one optical system to capture range and photometric data simultaneously. Therefore, incoming light travels the same optical path and a beam splitter is used for decomposition into photometric and range information. Due to this setup, we assume that a pair $(\tilde{\boldsymbol{u}}_z, \tilde{\boldsymbol{u}}_r)$ of corresponding pixels given in homogeneous coordinates is related by a homography. The mapping f used for sensor fusion is given by:

$$\tilde{\boldsymbol{u}}_z \cong \boldsymbol{H}_{rz} \tilde{\boldsymbol{u}}_r, \tag{1}$$

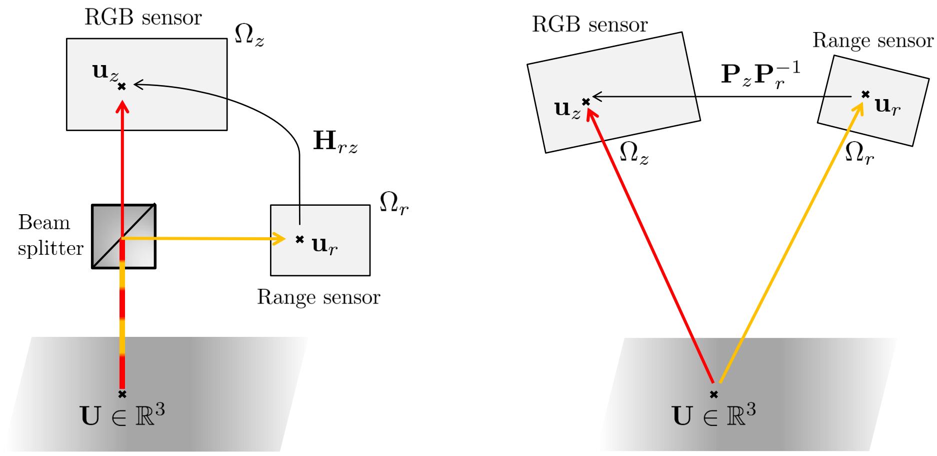


Figure 1: System setups investigated in this paper: In a beam splitter setup (left) range and photometric data is related to each other according to a homography. In a general stereo vision setup (right), sensor fusion is addressed by stereo calibration.

where $\boldsymbol{H}_{rz} \in \mathbb{R}^{3\times 3}$ denotes a homography to describe the transformation of the pixel coordinates (see Fig. 1a) and $\cong$ denotes equality up to a scale factor for homogeneous coordinates. The homography $\boldsymbol{H}_{rz}$ is calibrated by means of least-squares estimation based on point correspondences obtained from a checkerboard calibration pattern with self-encoded markers (Haase et al., 2013a). Once $\boldsymbol{H}_{rz}$ is estimated, photometric data is transformed pixel-wise into the domain of range images.

3.2. Stereo Vision

In the more general case of a stereo camera setup, the homography assumption does not hold true. In our work, we use a range sensor and an RGB camera to observe the scene from two different viewpoints (see Fig. 1b). Both sensors are displaced to each other with a small baseline to reduce occlusions. In order to estimate f, we propose a two-stage stereo calibration procedure:

1. An initial solution for f is estimated by extrinsic stereo calibration as follows: To obtain reliable depth information, we apply bilateral filtering, temporal averaging and defect pixel interpolation to raw range data $\boldsymbol{y}$ in a real-time preprocessing pipeline (Wasza et al., 2011a,b). Based on the preprocessed depth information $\tilde{\boldsymbol{y}}$, each point $\boldsymbol{u}_r$ in $\tilde{\boldsymbol{y}}$ is reprojected to its 3-D position $\tilde{\boldsymbol{U}}$ in homogeneous coordinates:
$$\tilde{\boldsymbol{U}} \cong \boldsymbol{P}_r^{-1} \begin{pmatrix} \boldsymbol{u}_r & \tilde{\boldsymbol{y}}(\boldsymbol{u}_r) & 1 \end{pmatrix}^\top, \qquad (2)$$
where $\boldsymbol{P}_r \in \mathbb{R}^{4\times 4}$ denotes the projection matrix to transform 3-D scene points to the image plane of the range sensor (Park et al., 2011). Then, $\tilde{\boldsymbol{U}}$ is projected to the RGB image plane to determine the associated pixel position $\tilde{\boldsymbol{u}}_z$:
$$\tilde{\boldsymbol{u}}_z \cong \boldsymbol{P}_z \boldsymbol{P}_r^{-1} \begin{pmatrix} \boldsymbol{u}_r & \tilde{\boldsymbol{y}}(\boldsymbol{u}_r) & 1 \end{pmatrix}^\top, \qquad (3)$$
where $\boldsymbol{P}_z \in \mathbb{R}^{3\times 4}$ is the projection matrix of the RGB sensor. The parameters of the projection matrices $\boldsymbol{P}_r$ and $\boldsymbol{P}_z$ are estimated by means of stereo calibration using a calibration pattern (Bouguet, 2013).
2. Finally, we perform a further refinement of the initial mapping to compensate for systematic errors in range images, e. g. intensity-dependent errors (Kolb et al., 2010). Therefore, we refine the translation vector of the extrinsic camera parameters in a grid search to maximize the normalized mutual information (Pluim et al., 2003) between the range image and the mapped photometric data.

4. Maximum A-Posteriori Framework

We reconstruct an HR image $\boldsymbol{x} \in \mathbb{R}^N$ from a set of LR range images $\boldsymbol{y}^{(1)}, \ldots, \boldsymbol{y}^{(K)}$, $\boldsymbol{y}^M \in \mathbb{R}^M$ that are warped to each other by means of MAP estimation. For this purpose, we exploit 2-D motion in the image plane caused by 3-D camera or object motion. The sampling of $\boldsymbol{x}$ is related to the sampling of each LR image by the magnification factor denoted by s, where $N = s^2 \cdot M$. The HR image $\boldsymbol{x}$

coincides with one LR frame $\boldsymbol{y}^{(r)}$, $1 \leq r \leq K$ used as reference in our motion model.

4.1. Generative Image Model

We use a generative image model (Elad and Feuer, 1997) to formulate the relation between an ideal HR image $\boldsymbol{x}$ and a LR frame $\boldsymbol{y}^{(k)}$ as follows: Each $\boldsymbol{y}^{(k)}$ is warped with respect to $\boldsymbol{x}$ according to a 2-D geometric transformation in the image plane. In addition, range values in successive frames corresponding to the same scene point are diverse due to general 3-D motion, which cannot be modeled by 2-D displacements, e. g. out-of-plane movements. Each frame is affected by blur induced by the *point spread function* (PSF) originating from the optical system and the sensor array. The blurred image is downsampled with respect to $\boldsymbol{x}$. Finally, each frame is disturbed by random noise induced by the sensor array and non-ideal optical components. The model to obtain $\boldsymbol{y}^{(k)}$ from $\boldsymbol{x}$ is given as:

$$\begin{aligned} \boldsymbol{y}^{(k)} &= \boldsymbol{G}_m^{(k)} \boldsymbol{D} \boldsymbol{B} \boldsymbol{M}^{(k)} \boldsymbol{x} + \boldsymbol{g}_a^{(k)} + \boldsymbol{\epsilon}^{(k)} \\ &= \boldsymbol{G}_m^{(k)} \boldsymbol{W}^{(k)} \boldsymbol{x} + \boldsymbol{g}_a^{(k)} + \boldsymbol{\epsilon}^{(k)}, \end{aligned} \tag{4}$$

where $\boldsymbol{M}^{(k)}$, $\boldsymbol{B}$ and $\boldsymbol{D}$ model subpixel motion, blur and downsampling, respectively. For convenience, these effects are combined into one system matrix $\boldsymbol{W}^{(k)} \in \mathbb{R}^{M \times N}$. $\boldsymbol{\epsilon}^{(k)} \in \mathbb{R}^M$ denotes additive and space invariant noise. $\boldsymbol{G}_m^{(k)} \in \mathbb{R}^{M \times M}$ and $\boldsymbol{g}_a^{(k)} \in \mathbb{R}^M$ models multiplicative and additive diversity of range values in successive frames caused, e. g. by out-of-plane movements. This transformation is mathematically equivalent to those proposed by Capel (2004) for modeling of photometric diversity of color images. $\boldsymbol{B}$ and $\boldsymbol{D}$ are assumed to be known, whereas $\boldsymbol{M}^{(k)}$ as well as $\boldsymbol{G}_m^{(k)}$ and $\boldsymbol{g}_a^{(k)}$ are estimated using image registration. For the reference frame $\boldsymbol{y}^{(r)}$, we set $\boldsymbol{M}^{(r)}$, $\boldsymbol{G}_m^{(r)}$ and $\boldsymbol{g}_a^{(r)}$ to the identity such that $\boldsymbol{x}$ coincides with $\boldsymbol{y}^{(r)}$.

If we assume a space invariant PSF given by the blur kernel $B(\boldsymbol{u})$, the system matrix is composed according to:

$$W_{mn} = B\left(||\boldsymbol{v}_n - \boldsymbol{u}'_m||_2\right), \tag{5}$$

where $\boldsymbol{v}_n \in \mathbb{R}^2$ are the coordinates of the n^{th} pixel in $\boldsymbol{x}$ and $\boldsymbol{u}'_m \in \mathbb{R}^2$ are the coordinates of the m^{th} pixel in $\boldsymbol{y}^{(k)}$ mapped onto the HR grid according to $\boldsymbol{M}^{(k)}$ (Tipping and Bishop, 2003). We normalize $\boldsymbol{W}^{(k)}$ such that $\sum_n W_{mn} = 1$.

4.2. MAP Estimator

In our work, we employ MAP estimation (Elad and Feuer, 1997) to reconstruct the HR image $\boldsymbol{x}$ from a sequence of LR frames $\boldsymbol{y}^{(1)}, \ldots, \boldsymbol{y}^{(K)}$. The objective function requires a data term and a regularizer for $\boldsymbol{x}$ according to:

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \left\{ \Gamma_{\text{data}}(\boldsymbol{x}) + \lambda \Gamma_{\text{reg}}(\boldsymbol{x}) \right\}. \tag{6}$$

The data term $\Gamma_{\text{data}}(\boldsymbol{x})$ measures the fidelity of an estimate $\boldsymbol{x}$ with respect to the observed LR range data $\boldsymbol{y}^{(1)}, \ldots, \boldsymbol{y}^{(K)}$. If $\boldsymbol{\epsilon}^{(k)}$ is assumed to be additive, zero-mean and space invariant Gaussian noise, the data term is given by:

$$\Gamma_{\text{data}}(\boldsymbol{x}) = \sum_{k=1}^{K} \left| \left| \boldsymbol{y}^{(k)} - \boldsymbol{G}_m^{(k)} \boldsymbol{W}^{(k)} \boldsymbol{x} - \boldsymbol{g}_a^{(k)} \right| \right|_2^2. \tag{7}$$

Please see Sect. 5 for details on the design of the data term in our approach.

Since super-resolution is known to be an ill-posed problem, prior knowledge is incorporated into the reconstruction process using a regularizer $\Gamma_{\text{reg}}(\boldsymbol{x})$ weighted by $\lambda \geq 0$. This guides the estimation to reliable solutions $\hat{\boldsymbol{x}}$. The specific regularizer employed in our framework is introduced in Sect. 5.3.

5. Multi-Sensor Range Super-Resolution

Next, we introduce our proposed multi-sensor super-resolution framework, where super-resolution of one modality is guided by a complementary modality. In terms of RI, photometric information is utilized as guidance for range super-resolution. Depending on the system setup, we choose one of the sensor data fusion approaches presented in Sect. 3 to align both image types. In our framework, motion estimation is realized by optical flow computation on high-quality photometric data. Then, photometric displacement fields are transferred to range images to obtain accurate displacements of subpixel accuracy for super-resolution reconstruction. Since 2-D displacement fields cannot model 3-D motion in general, an additional range correction scheme is introduced. Super-resolution is formulated as MAP estimation guided by a novel adaptive regularizer that exploits photometric data for improved edge reconstruction. See Fig. 2 for a flowchart of the proposed method.

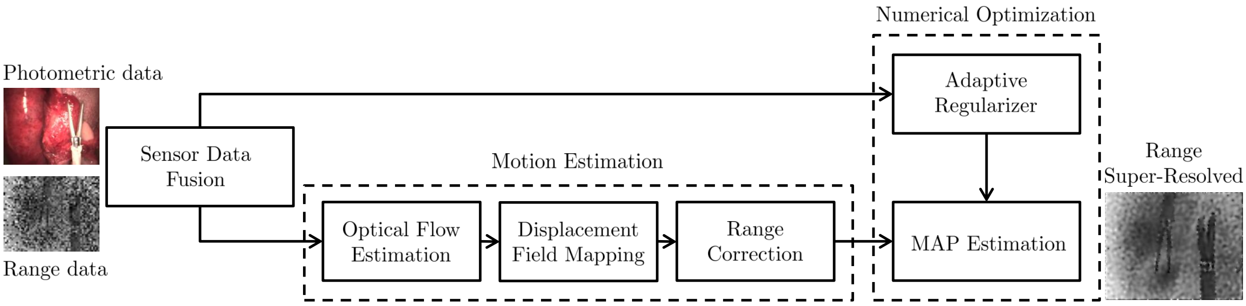


Figure 2: Flowchart of the proposed method to obtain a super-resolved range image from multiple LR frames: We exploit sensor fusion of range data with complementary photometric information. For motion estimation, displacement fields are obtained by optical flow on photometric data and transferred to the associated range images. Range correction compensates diverse depth values due to general 3-D camera motion (e. g. out-of-plane movements) not modeled by optical flow. Numerical optimization is implemented as maximum a-posteriori (MAP) estimation guided by adaptive regularization based on photometric information.

5.1. Geometric Registration for Motion Estimation

We use non-rigid displacement fields estimated by means of optical flow to model $\boldsymbol{M}^{(k)}$ in the underlying system matrix defined by Eq. (5). Zhao and Sawhney (2002) have suggested that this is feasible for small noise and accurate flow estimation resulting in small warping errors. For these reasons, optical flow is computed on high-quality photometric data to obtain displacement fields of subpixel accuracy. This avoids direct flow estimation on LR range data. We determine photometric displacement fields $\boldsymbol{w}_z^{(k)} : \Omega_z \rightarrow \mathbb{R}^2$, $\boldsymbol{w}_z^{(k)}(\boldsymbol{u}_z) = (\boldsymbol{w}_{z,1}^{(k)}(\boldsymbol{u}_z), \boldsymbol{w}_{z,2}^{(k)}(\boldsymbol{u}_z))^\top$ for $\boldsymbol{z}^{(k)}$ with respect to the reference $\boldsymbol{z}^{(r)}$ to transform each point $\boldsymbol{u}_z$ from $\boldsymbol{z}^{(k)}$ to its position $\boldsymbol{u}'_z$ in $\boldsymbol{z}^{(r)}$ according to:

$$\boldsymbol{u}'_z = \boldsymbol{u}_z + \boldsymbol{w}_z^{(k)}(\boldsymbol{u}_z). \tag{8}$$

The central frame $\boldsymbol{z}^{(r)}$, $r = \lceil K/2 \rceil$ is chosen as reference to avoid error propagation in case of misregistration for individual frames. For optical flow computation, we use a variational approach solved in a coarse-to-fine scheme (Liu, 2009).

Based on sensor data fusion, photometric displacements $\boldsymbol{w}_z$ are transferred to associated range displacements $\boldsymbol{w}_r : \Omega_r \rightarrow \mathbb{R}^2$ according to a mapping $\Delta : \mathbb{R}^2 \rightarrow \mathbb{R}^2$. Therefore, it is assumed that each $\boldsymbol{y}^{(k)}$ is fused with the associated photometric data $\boldsymbol{z}^{(k)}$ using one of the techniques presented in Sect. 3. Then, the mapping of photometric displacement fields is realized as follows:

1. Intermediate displacements $\tilde{\boldsymbol{w}}_r : \Omega_z \rightarrow \mathbb{R}^2$ are calculated by component-wise rescaling of $\boldsymbol{w}_z$ with the relative resolution $\boldsymbol{l} = (l_1, l_2)^\top, l_i = M_i/L_i$ denoting the ratio of image sizes between range data (size $M_1 \times M_2$) and photometric data (size $L_1 \times L_2$).
2. Finally, $\boldsymbol{w}_r$ is obtained by resampling $\tilde{\boldsymbol{w}}_r$ to the coordinate grid Ω_r expressed as:

$$\boldsymbol{w}_r(\boldsymbol{u}_r) = \Delta \begin{pmatrix} l_1 \cdot \boldsymbol{w}_{z,1}(\boldsymbol{u}_z) \\ l_2 \cdot \boldsymbol{w}_{z,2}(\boldsymbol{u}_z) \end{pmatrix}. \tag{9}$$

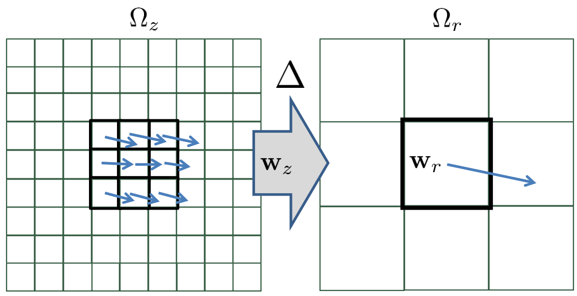


Figure 3: Range displacements $\boldsymbol{w}_r$ (right) are determined from associated photometric displacements $\boldsymbol{w}_z$ (left) using rescaling and resampling by Δ. We employ the element-wise median for resampling.

In order to obtain accurate, denoised displacement fields $\boldsymbol{w}_r$ while preserving motion discontinuities, the resampling operator Δ is implemented by element-wise median filtering. Resampling is performed for the associated photometric displacements $\boldsymbol{w}_z$ in each coordinate direction in a patchwise scheme as shown in Fig. 3.

5.2. Range Data Correction

If we allow general 3-D motion for successive frames, this cannot be modeled by 2-D optical flow estimation in general. This becomes obvious in case

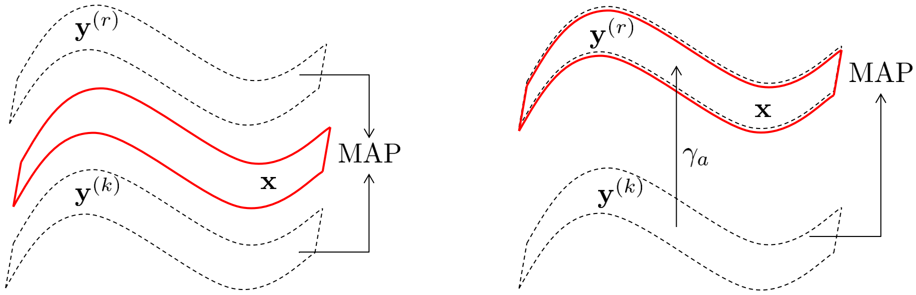


Figure 4: Range correction in case of out-of-plane translation between two frames $\boldsymbol{y}^{(r)}$ and $\boldsymbol{y}^{(k)}$ (black, dashed lines) aligned by optical flow: Without range correction, the super-resolved image $\boldsymbol{x}$ (red, solid line) does not render a true 3-D surface (left). Range correction compensates out-of-plane motion (offset γ_a) resulting in the desired surface $\boldsymbol{x}$ aligned to $\boldsymbol{y}^{(r)}$ (right).

of non-planar camera motion orthogonal to the image plane, i. e. out-of-plane motion. Neglecting this effect, super-resolution fails as demonstrated in Fig. 4. This issue may be compared to fusing multiple intensity images having photometric differences. To take this effect into account for range super-resolution, we introduce $\boldsymbol{g}_a$ and $\boldsymbol{G}_m$ in our model defined in Eq. (4).

One crucial type of motion is out-of-plane translation (see Fig. 4). This is modeled by $\boldsymbol{g}_a = \gamma_a \mathbf{1}_M$ for the all-ones vector $\mathbf{1}_M$ and an offset $\gamma_a \in \mathbb{R}$. Additionally, a tilting of the camera relative to the scene or projective distortions are modeled by $\boldsymbol{G}_m = \gamma_m \boldsymbol{I}_M$ for positive $\gamma_m \in \mathbb{R}$, where $\boldsymbol{I}_M$ denotes the $M \times M$ identity matrix. We neglect higher-order effects such as non-rigid deformations in out-of-plane direction and use the affine model:

$$y' = \gamma_m y + \gamma_a, \tag{10}$$

where $y \in \mathbb{R}$ and $y' \in \mathbb{R}$ are corresponding range values in reference $\boldsymbol{y}^{(r)}$ and frame $\boldsymbol{y}^{(k)}$, $k \neq r$ after 2-D geometric alignment according to Eq. (8). The estimation of γ_m and γ_a is formulated as line fitting problem for given samples $\mathcal{Y} = \{(y_m, y'_m)\}$. However, robust parameter estimation techniques are required due to the perturbation of range data by random noise and outliers caused, e. g. by systematic errors or inaccurate geometric alignments.

We adopt a photometric registration scheme of intensity images (Capel, 2004) for implicit correction of diverse range values. First, a 3×3 median filter is applied to $\boldsymbol{y}^{(r)}$ and $\boldsymbol{y}^{(k)}$ for noise reduction. Then, the robust *M-estimator sample consensus* (MSAC) algorithm (Torr and Zisserman, 2000) is used. In contrast to conventional RANSAC, the costs assigned to each inlier are proportional to its data likelihood instead of a constant value to make the threshold selection for inlier classification less crucial. The objective function optimized by MSAC in a probabilistic manner is given by:

$$\begin{aligned} (\hat{\gamma}_m, \hat{\gamma}_a) &= \arg \min_{\gamma_m, \gamma_a} \sum_i \rho((y' - \gamma_m y - \gamma_a)^2) \\ \rho(d^2) &= \begin{cases} d^2 & \text{if } d^2 < T^2 \\ T^2 & \text{otherwise} \end{cases} . \end{aligned} \tag{11}$$

The threshold T is set to $T = 1.96\sigma_n$ such that 5% of the inliers are suspected as outliers if we assume range values disturbed by zero-mean Gaussian noise with standard deviation σ_n. Finally, the set of inliers $\tilde{\mathcal{Y}} = \{(\tilde{y}_m, \tilde{y}'_m) \,|m \text{ is inlier}\} \subseteq \mathcal{Y}$ having minimal costs according to Eq. (11) is used to obtain a refined solution for γ_m and γ_a using linear least-squares optimization.

5.3. Regularization

Super-resolution is known to be an ill-posed problem (Baker and Kanade, 2002; Lin and Shum, 2004) and thus regularization is required to obtain reliable solutions. Various regularization techniques have been proposed in literature to incorporate prior knowledge into the image reconstruction process. Common priors exploit smoothness, piecewise smoothness or sparsity of images in gradient or transform domains. However, these assumptions often ignore discontinuities. For this reason, we propose a spatially adaptive regularizer plugged into Eq. (6) for MAP estimation. Our regularization term exploits range data as well as complementary photometric information and is given by:

$$\Gamma_{reg}(\boldsymbol{x}) = \alpha(\boldsymbol{z}, \boldsymbol{x})^\top \phi(\boldsymbol{H}\boldsymbol{x}), \tag{12}$$

where $\alpha : \Omega_r \times \Omega_z \rightarrow \mathbb{R}^N$ are pixel-wise spatially adaptive weights derived below and $\phi(\tilde{\boldsymbol{x}})$ denotes a loss function applied to a high-pass filtered version of $\boldsymbol{x}$ given as $\tilde{\boldsymbol{x}} = \boldsymbol{H}\boldsymbol{x}$. Due to the weights, $\Gamma_{reg}(\boldsymbol{x})$ is now parametrized by complementary photometric data $\boldsymbol{z}$ fused with the HR estimate $\boldsymbol{x}$. For $\phi(\tilde{\boldsymbol{x}})$, we use a pseudo Huber loss function given by:

$$\phi(\tilde{\boldsymbol{x}}) = \sum_{i=1}^{N} \tau \left(\sqrt{1 + \left(\frac{\tilde{x}_i}{\tau} \right)^2} - 1 \right). \tag{13}$$

The advantage of using the Huber loss function compared to Tikhonov regularization is that edges are better preserved. Unlike the L_1 norm, the Huber function has continuous derivatives. For $\boldsymbol{H}$, we choose a Laplacian such that:

$$\boldsymbol{H}\boldsymbol{x} \equiv \boldsymbol{X} * \tilde{\boldsymbol{H}}, \tag{14}$$

where the Laplacian filter kernel $\tilde{\boldsymbol{H}}$ is given as:

$$\tilde{\boldsymbol{H}} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{pmatrix}. \tag{15}$$

5.3.1. Adaptive Regularization

Our basic idea is to decrease the impact of the regularizer on edges compared to flat regions to improve edge reconstruction. Recently, spatially weighted regularization has been proposed working solely on LR data based on edge detection (Yuan et al., 2012). In order to improve robustness, we exploit photometric data based on the assumption that discontinuities in both domains are correlated.

Spatially adaptive weights $\alpha(\boldsymbol{z}, \boldsymbol{x})$ are derived as follows from super-resolved range data $\boldsymbol{x}$ and fused photometric information $\boldsymbol{z}$. First, we perform edge detection on photometric data $\boldsymbol{z}$ to obtain a binary edge map $\tau(\boldsymbol{u}_z)$, where $\tau(\boldsymbol{u}_z) = 1$ for edge points. While edges in $\boldsymbol{x}$ are usually also visible in $\boldsymbol{z}$ due to scene illumination, vice versa this assumption may be violated, e. g. for texture. To avoid transfer of false edge information, we employ a similarity measure $D : \Omega_z \times \Omega_r \rightarrow \mathbb{R}$ to weight the relationship between $\boldsymbol{x}$ and $\boldsymbol{z}$ in a pixel-wise manner. First, $\boldsymbol{x}$ is resampled to the dimension of $\boldsymbol{z}$. Then, the spatially adaptive weights are computed pixel-wise according to:

$$\alpha(\boldsymbol{z}, \boldsymbol{x}) = \begin{cases} \exp\left(-\frac{D(\boldsymbol{u}_z, \boldsymbol{u}_r)}{\tau_0}\right) & \tau(\boldsymbol{u}_z) = 1 \\ 1 & \text{otherwise} \end{cases}. \tag{16}$$

where $\tau_0 \in \mathbb{R}$, $\tau_0 > 0$ denotes a contrast factor, and $(\boldsymbol{u}_z, \boldsymbol{u}_r)$ are corresponding pixels in $\boldsymbol{z}$ and $\boldsymbol{x}$, respectively. Thus, we assign to $(\boldsymbol{u}_z, \boldsymbol{u}_r)$ a smaller weight if $\boldsymbol{u}_z$ in $\boldsymbol{z}$ is an edge point and the similarity $D(\boldsymbol{u}_z, \boldsymbol{u}_r)$ is high. Opposed to that, a constant weight λ is used in flat regions.

5.3.2. Similarity Measure

We employ mutual information as multi-modal similarity measure for $D(\boldsymbol{u}_z, \boldsymbol{u}_r)$. Since pixel-wise analysis is required, the local mutual information (LMI) (Pluim et al., 2003) is used, which is given by:

$$I(\boldsymbol{u}_z, \boldsymbol{u}_r) = \sum_{\boldsymbol{v}_z, \boldsymbol{v}_r \in \mathcal{N}} P(\boldsymbol{v}_z, \boldsymbol{v}_r) \log \left(\frac{P(\boldsymbol{v}_z, \boldsymbol{v}_r)}{P(\boldsymbol{v}_z) P(\boldsymbol{v}_r)} \right), \tag{17}$$

where $P(\boldsymbol{v}_z, \boldsymbol{v}_r)$ is the joint probability distribution for range and photometric data, and $P(\boldsymbol{v}_z)$ and $P(\boldsymbol{v}_r)$ are the marginals, respectively. LMI is calculated pixel-wise for an image patch composed from the $p \times p$ neighborhood $\mathcal{N}$ centered at $(\boldsymbol{u}_z, \boldsymbol{u}_r)$. We use a normalized variant of LMI according to:

$$D(\boldsymbol{u}_z, \boldsymbol{u}_r) = -\frac{I(\boldsymbol{u}_z, \boldsymbol{u}_r)}{\sum_{\boldsymbol{v}_z, \boldsymbol{v}_r \in \mathcal{N}} P(\boldsymbol{v}_z, \boldsymbol{v}_r) \log \left(P(\boldsymbol{v}_z, \boldsymbol{v}_r) \right)} \tag{18}$$

This is not the first time that LMI is employed for image improvement. Guo and Huang (2008) have proposed LMI adaptive total variation denoising. We adopt this idea to exploit statistical dependency between range images fused with photometric data.

5.4. Numerical Optimization

In summary, multi-sensor super-resolution is implemented as minimization of the objective function:

$$\begin{aligned} \Gamma(\boldsymbol{x}) = & \sum_{k=1}^{K} \left| \left| \boldsymbol{y}^{(k)} - \gamma_m^{(k)} \boldsymbol{W}^{(k)} \boldsymbol{x} - \gamma_a^{(k)} \right| \right|_2^2 \\ & + \lambda \cdot \alpha(\boldsymbol{z}, \boldsymbol{x})^\top \phi(\boldsymbol{H}\boldsymbol{x}). \end{aligned} \tag{19}$$

This is a non-convex problem and direct optimization using gradient-based techniques is hard to perform as the adaptive weights are not given in a closed-form solution. As extension of our approach introduced in (Köhler et al., 2013), we propose an efficient two-stage optimization procedure. In the first stage, we initialize the weights by $\alpha(\boldsymbol{z}, \boldsymbol{x}) = \boldsymbol{1}_N$, resulting in multi-sensor super-resolution without adaptive regularization. Then, if the weights are kept fixed, the gradient of Eq. (19) is given by:

$$\nabla_{\boldsymbol{x}}\Gamma(\boldsymbol{x}) = \lambda \cdot \alpha(\boldsymbol{z}, \boldsymbol{x})^\top \boldsymbol{H}^\top \frac{\partial \phi}{\boldsymbol{x}}(\boldsymbol{H}\boldsymbol{x}) - 2\sum_{k=1}^{K} \gamma_m^{(k)} \boldsymbol{W}^{(k)\top} \left(\boldsymbol{y}^{(k)} - \gamma_m^{(k)} \boldsymbol{W}^{(k)} \boldsymbol{x} - \gamma_a^{(k)} \right) \tag{20}$$

For numerical minimization of the resulting convex optimization problem, we employ an iterative Scaled Conjugate Gradients (SCG) scheme (Nabney, 2002). An initial guess for SCG is obtained by bicubic upsampling of the reference frame $\boldsymbol{y}^{(r)}$. For further refinement in a second stage, spatially adaptive weights $\boldsymbol{\alpha}$ based on $\boldsymbol{x}_1$ and fused photometric data $\boldsymbol{z}$ are derived according to Eq. (16) yielding an adaptive version of our method. For increased robustness, we take the temporal median of $\boldsymbol{z}^{(1)}, \ldots, \boldsymbol{z}^{(K)}$ for $\boldsymbol{z}$ to derive spatially adaptive weights. Then, SCG is performed for a second time based on the weights $\boldsymbol{\alpha}$ and $\boldsymbol{x}_1$ as initial guess, which yields the final super-resolved image $\boldsymbol{x}_2$.

This procedure is a particular case of coordinate-wise descent with $S = 2$ iteration stages. One may also employ $S > 2$ iterations to update adaptive weights gradually. However, we found that improvements becomes marginal while the computational effort is increased.

6. Experiments and Results

Our experiments are divided into two parts. Performance and robustness of our framework are quantitatively evaluated for synthetic images. In the second part, we address image-guided surgery in 3-D endoscopy and open surgery and present results for ex-vivo data. Supplementary material including a Matlab implementation of our framework and evaluation datasets is available on our webpage[1].

6.1. Synthetic Images

Synthetic range images were generated based on ground truth data. We used a RI simulator (Wasza et al., 2011a) to obtain photometric data fused with range data of a laparoscopic scene designed in collaboration with a medical expert (see Fig 5).

[1] See http://www5.cs.fau.de/research/software and http://www5.cs.fau.de/en/research/data

6.1.1. Experimental Setting

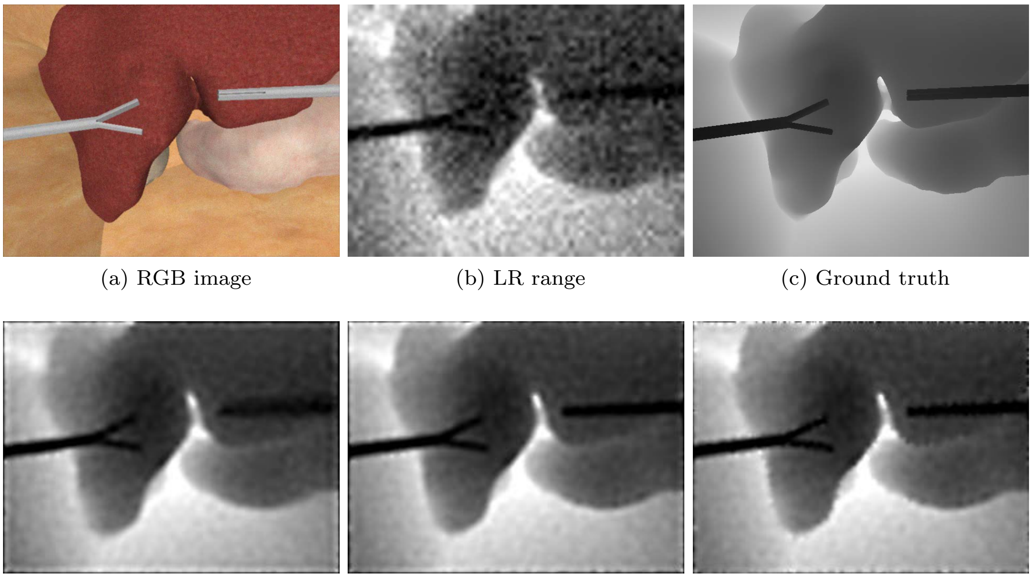
We simulated the underlying conditions of 3-D endoscopy. As resolution we chose 640×480 px for photometric data and 64×48 px for range images reflecting the setup of the hybrid 3-D endoscope used in ex-vivo experiments below. Photometric data was given as color image in RGB space. In order to simulate a real RGB sensor, each color image was blurred with Gaussian filter (5×5 pixel) and disturbed by additive, zero-mean Gaussian noise ($\sigma_n = 0.001$). Range images were a downsampled version of ground truth data and disturbed by a Gaussian PSF ($\sigma_b = 0.5$) as well as zero-mean Gaussian noise ($\sigma_n = 0.05$). Please note that we did not consider hardware-dependent, systematic errors in our general study opposed to specific simulations as done, e. g. for ToF (Maier-Hein et al., 2010).

We generated 12 data sets (S1 - S12) each showing the scene from a different perspective and with different objects such as endoscopic tools. Three types of camera motion were analyzed. Small planar motion was simulated for S1 - S4. For S5 - S8, moderate out-of-plane motion compared to the measurement range was simulated. For S9 - S12, out-of-plane motion was substantially increased. Camera motion was simulated by a random rigid transformation in 3-D space. Additionally, we simulated independent movements of tools and organs in all data sets.

6.1.2. Evaluated Methods

We compared three methods in our experiments:

1. As baseline, single-sensor super-resolution (termed as SSR) (Schuon et al., 2009) was considered. Here, photometric information is not taken into account and motion estimation is performed on range data directly. For a fair comparison, we also make use of our range correction scheme. Optical flow estimation and regularization based on the Huber function is done analogously to our method.
2. In our multi-sensor approach (termed as MSR) (Köhler et al., 2013), we derive subpixel displacements from photometric data. This corresponds to the outcome of the first stage in our two-stage optimization proposed in Sect. 5.4.
3. For an explicit evaluation of adaptive regularization, we considered adaptive multi-sensor super-resolution (termed as AMSR) as extension of MSR. The result of AMSR is the final outcome of the two-stage optimization.



(a) RGB image (b) LR range (c) Ground truth

(d) SSR (e) MSR (f) AMSR

Figure 5: Results obtained from synthetic data (magnification $s = 4$, $K = 31$ frames) for the different approaches: RGB data (a), LR range data (b), our ground truth (c) and the outcomes of SSR (d) as well as the proposed MSR (e) and AMSR (f).

For MSR as well AMSR, we computed the optical flow in RGB space. Edge detection required for adaptive regularization was performed using the Sobel operator on the V-channel of color images transformed to the HSV space. Super-resolution was evaluated for different magnification factors and number of LR frames. For magnification $s = 2$, we used $K = 7$ frames, whereas for $s = 4$ we chose $K = 31$. All approaches were implemented in a sliding window scheme, where K successive frames in a data set were utilized to reconstruct a super-resolved image. Quantitative assessment was done by comparing a super-resolved image to ground truth data using the full-reference quality measures peak-signal-to-noise ratio (PSNR) and structural similarity (SSIM) (Wang et al., 2004).

6.1.3. Parameter Setting and Optimization

Throughout our experiments, SCG was applied with termination tolerance 10^{-3} and the maximum iteration number was set to 50. For regularization using the Huber function, we set $\tau = 5 \cdot 10^{-3}$ for all evaluated methods[2]. For the selection of an appropriate regularization weight λ, we performed a grid search based on a single data set. This was done for the SSR and MSR approach, where a single weight is required. See Fig. 6 for the impact of λ to PSNR and SSIM measures. If λ is set too small, super-resolved data is affected by residual noise, whereas a large value leads to oversmooth solutions. The impact of λ was independent of the super-resolution approach and we set $\lambda = 0.4$ for further experiments.

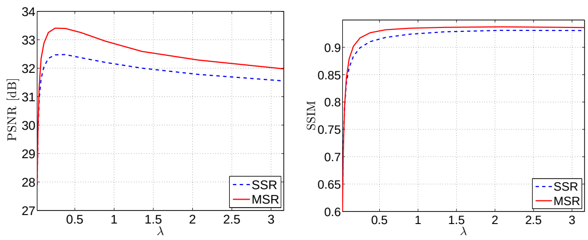
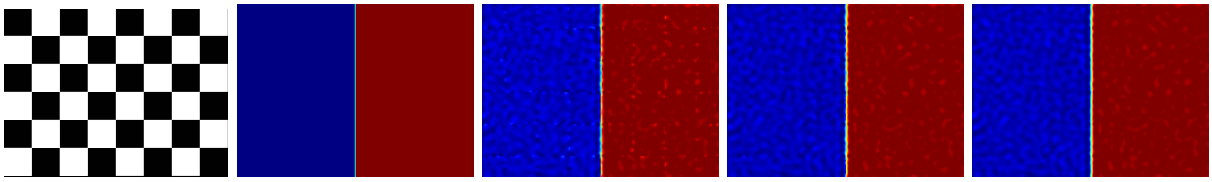


Figure 6: Impact of regularization weight λ to super-resolved data obtained by single-sensor (SSR) and multi-sensor super-resolution (MSR) on one data set.

For AMSR, we set the size of neighborhood $\mathcal{N}$ to 25×25 pixels to compute LMI. In this experiment, an artificial checkerboard scene was used as worst-case scenario for adaptive regularization since most edges of the checkerboard visible in photometric data did not correspond with depth discontinuities (see Fig. 7). We investigated the impact of the

[2] One can also use automatic parameter selection (Nguyen et al., 2001), which is beyond the scope of this work.



(a) RGB (b) Ground truth (c) AMSR ($\tau_0 = 0.005$) (d) AMSR ($\tau_0 = 0.025$) (e) AMSR ($\tau_0 = 0.075$)

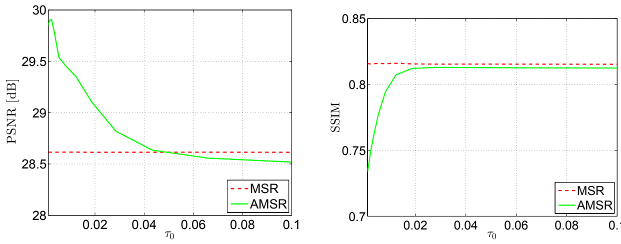


Figure 7: Checkerboard to adjust the contrast factor τ_0 for adaptive multi-sensor super-resolution (AMSR): RGB data (a), ground truth (b) and outcomes of AMSR for increasing τ_0 (c) - (e). Please see the checkerboard texture transferred to super-resolved data for $\tau_0 = 0.005$ but suppressed for higher τ_0. Peak-signal-to-noise ratio (PSNR) and structural similarity (SSIM) measures were averaged over ten sequences.

contrast factor τ_0 to PSNR and SSIM measures. If τ_0 is chosen too large, AMSR converges to MSR and is not longer adaptive. In case of too small τ_0, photometric texture is treated as range edge as shown in Fig. 7. For further experiments, we chose $\tau_0 = 0.025$ as appropriate value.

6.1.4. Experiment 1: Motion Estimation

The proposed MSR approach was compared to conventional SSR. For an explicit evaluation of motion estimation derived from photometric displacements, adaptive regularization was omitted in this experiment. The outcome of super-resolution was assessed for ten successive sub-sequences per data set using sliding window processing. Boxplots for PSNR and SSIM showing median values as well as the 25^{th} and 75^{th} percentile based on ten samples each are shown in Fig. 8. Mean and standard deviation over all data sets are reported in Tab. 1. On average, the proposed MSR approach improved PSNR (SSIM) by 1.8 dB (0.02) with respect to SSR for magnification $s = 4$. In case of $s = 2$ we obtained smaller changes for both error measures.

6.1.5. Experiment 2: Adaptive Regularization

We repeated our first experiment for comparison of MSR and AMSR to assess the benefit of adaptive regularization. The associated boxplots for both error measures are shown in Fig. 8. Please see Tab. 1 for comparison of mean and standard deviation over all data sets. For magnification $s = 4$, the PSNR (SSIM) measure was improved by 0.2 dB (0.01) if our adaptive regularizer was used. In total, we observed improved results by 2 dB (0.03) with respect to conventional SSR.

6.1.6. Experiment 3: Range Correction

For an explicit evaluation of range correction, its impact on the accuracy of super-resolved data was studied. For this purpose, AMSR was evaluated with and without the proposed correction scheme. Since diverse range values mainly cause a bias in super-resolved images whereas structural information is rendered correctly, PSNR was employed to assess super-resolution. See Fig. 9 for the corresponding boxplots. For mean and standard deviation over all data sets if range correction is not used see Tab. 1 (third and fourth row). In case of both magnification factors, the mean values were decreased and the standard deviations were increased if range correction was not employed.

6.1.7. Significance Tests

Statistical tests for differences between SSR, MSR and AMSR in terms of the error measures were performed using linear mixed-effects models. The dependency of observations within successive sub-sequences due to sliding window processing was accounted for by specification of autoregressive correlation structures. Models for both measures were

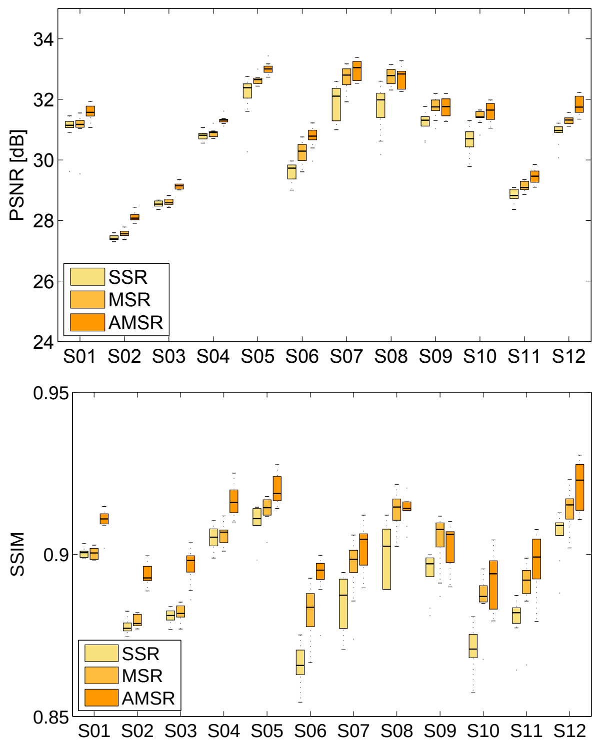


Figure 8: Peak-signal-to-noise ratio (PSNR) and structural similarity (SSIM) evaluated for single-sensor (SSR), multi-sensor (MSR) and adaptive multi-sensor super-resolution (AMSR). Each boxplot S1 to S12 was created for ten image sequences per data set. We calculated the median as well as the 25^{th} and 75^{th} percentile of PSNR (top row) and SSIM (bottom row) for magnification $s = 2$ (first column) and $s = 4$ (second column).

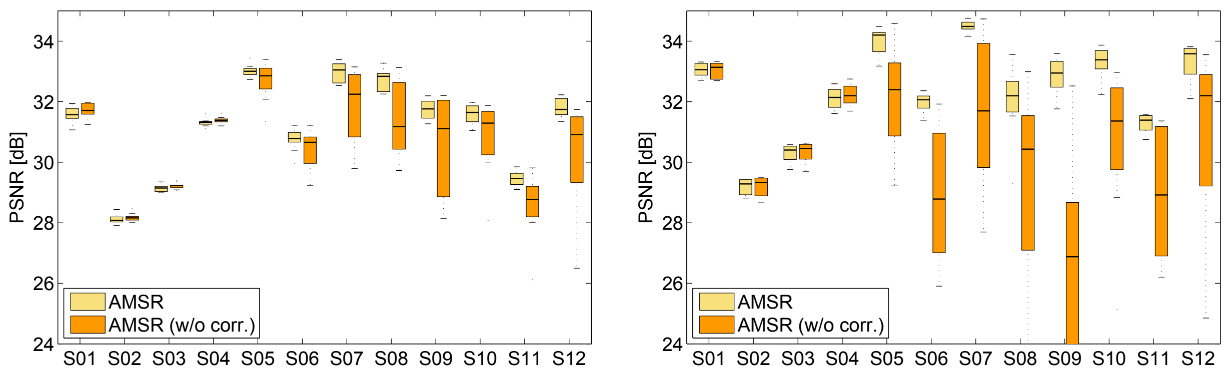


Figure 9: Comparison of adaptive multi-sensor super-resolution (AMSR) with and without range correction to evaluate the proposed correction scheme: Each boxplot S1 to S12 was created for ten image sequences per data set. We calculated the median as well as the 25^{th} and 75^{th} percentile of the peak-signal-to-noise ratio (PSNR) for magnification $s = 2$ (first column) and $s = 4$ (second column).

fitted in R version 3.1.1 applying the *nlme*-package (Pinheiro et al., 2014). P-values concerning differences between the approaches of ≤ 0.05 were considered to be statistically significant. For magnification $s = 4$, the PSNR was improved by 1.9 dB (95% confidence interval (CI): 0.6 - 3.1 dB, $P = 0.004$) for the MSR approach and 2.1 dB (95%-CI: 0.9 - 3.4 dB, $P = 0.001$) for the AMSR approach with respect to SSR. Similarly, SSIM was improved by 0.02 (95%-CI: 0.01 - 0.03, $P < 0.001$) for MSR and

Table 1: Mean and standard deviation of peak-signal-to-noise ratio (PSNR) and structural similarity (SSIM) over all data sets measured for the different super-resolution approaches. Each method was applied with (+, first and second row) and without the proposed range correction scheme (-, third and fourth row) for magnification $s = 2$ and $s = 4$, respectively.

Range correction	**Measure**	**Magnification $s = 2$**			**Magnification $s = 4$**		
		SSR	**MSR**	**AMSR**	**SSR**	**MSR**	**AMSR**
+	PSNR	30.4 ± 1.48	30.8 ± 1.63	**31.2 ± 1.52**	30.3 ± 1.69	32.1 ± 1.58	**32.3 ± 1.54**
+	SSIM	0.88 ± 0.03	0.90 ± 0.02	**0.90 ± 0.01**	0.89 ± 0.02	0.91 ± 0.01	**0.92 ± 0.01**
-	PSNR	30.0 ± 1.63	30.3 ± 1.69	30.6 ± 1.65	29.1 ± 2.83	30.1 ± 2.62	30.3 ± 2.67
-	SSIM	0.89 ± 0.02	0.90 ± 0.02	0.90 ± 0.01	0.89 ± 0.02	0.90 ± 0.02	0.91 ± 0.02

0.03 (95%-CI: 0.01 - 0.04, $P < 0.001$) for AMSR compared to SSR.

The differences between MSR and AMSR of 0.3 dB (95%-CI: -0.9 - 1.5 dB, $P = 0.652$) in case of PSNR and 0.005 (95%-CI: -0.005 - 0.02, $P = 0.328$) in case of SSIM were considerably smaller and not statistically significant. However, for $\approx 90\%$ of our test sequences, the errors measures were considerably improved by AMSR compared to MSR.

6.1.8. Robustness Analysis

Resolution of Photometric Data. Throughout our experiments, the resolution of photometric data was chosen ten times higher (640×480 px) than for associated range images (64×48 px). In practice there are limitations even for photometric data due to economical or technological constraints. We studied the impact of photometric resolution to the outcome of our framework. Therefore, we rescaled photometric data from its original resolution to smaller ones. The averaged error measures over ten sequences versus the scale of photometric data are plotted in Fig. 10. Our MSR approach outperformed SSR even if photometric data of low resolution is used. In case of AMSR, we observed decreased performance for Scale < 0.6 corresponding to a resolution of 384×288 px.

Disturbances in Photometric Data. In surgery, the reliability of photometric data may suffer from disturbances such as non-ambient lightning conditions, smoke or blood, which affects motion estimation and adaptive regularization. A frequently occurring issue are specular reflections, which makes optical flow estimation challenging. We studied the robustness of our framework in presence of reflections on organ surfaces based on three synthetic data sets R1 - R3. The associated boxplots for PSNR and SSIM for ten successive image sequences

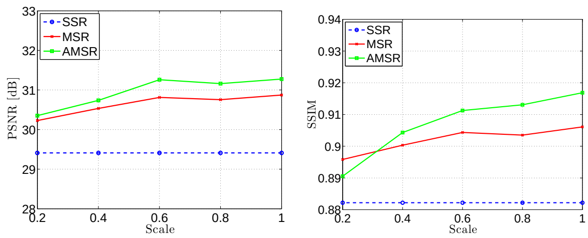
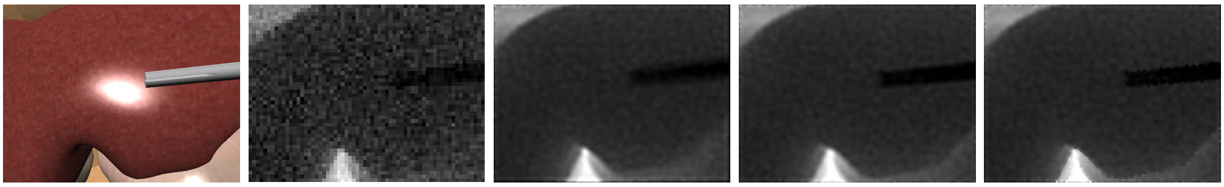


Figure 10: Robustness of multi-sensor (MSR) and adaptive multi-sensor super-resolution (AMSR) with respect to the resolution of photometric data scaled to different image sizes relative to the original resolution. Single-sensor super-resolution (SSR) is considered as baseline. The peak-signal-to-noise ratio (PSNR) and structural similarity (SSIM) measures were averaged over ten sequences.

per data set are shown in Fig. 11. The MSR as well as the AMSR method yield improved quality of super-resolved data in each set even in the presence of reflections.

Inaccuracy of Sensor Fusion. In all experiments reported above, perfect sensor data fusion was assumed. However, in practice the accuracy is limited due to calibration errors. To assess the robustness of our framework, we simulated misalignments between range and photometric data. Therefore, photometric data was shifted by a translation of random direction and different lengths ϵ_t relative to a perfect alignment. We increased ϵ_t measured in units of RGB pixels gradually and assessed the outcome of super-resolution. The error measures averaged over ten sequences are reported in Fig. 12 for magnification $s = 4$. We observed a minor decrease of ~ 0.5 dB for PSNR achieved by our methods even for large sensor fusion errors. In case of more severe errors ($\epsilon_t > 5$ px), AMSR was more affected compared to the MSR approach.



(a) RGB image (b) LR range (c) SSR (d) MSR (e) AMSR

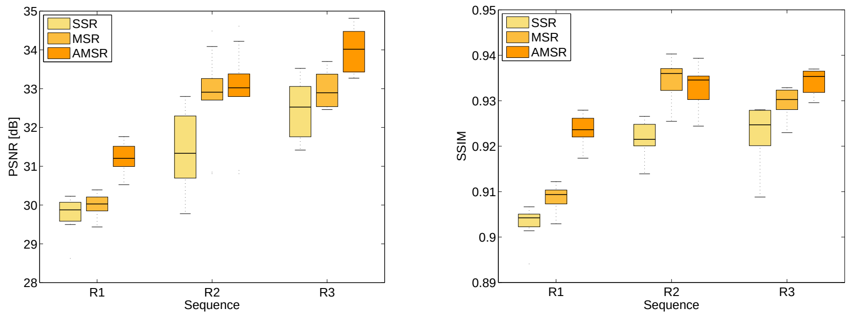


Figure 11: Results for synthetic data in presence of specular reflections (magnification $s = 4$, $K = 31$ frames): RGB data (a), LR range data (b) and the outcomes of single-sensor super-resolution (SSR) (c) as well as the proposed multi-sensor (MSR) (d) and adaptive multi-sensor (AMSR) method (e). The associated boxplots for the peak-signal-to-noise ratio (PSNR) and structural similarity (SSIM) were calculated for data sets R1 to R3 based on ten images sequences each.

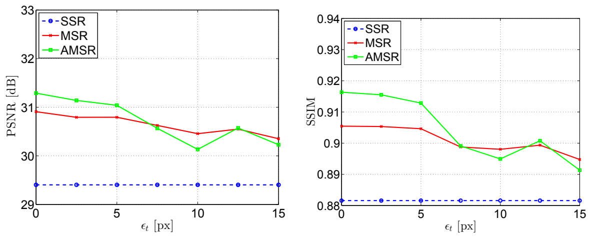


Figure 12: Robustness of multi-sensor (MSR) and adaptive multi-sensor super-resolution (AMSR) with respect to uncertainty of sensor fusion modeled by a random translation ϵ_t for photometric data relative to an ideal alignment. Single-sensor super-resolution (SSR) is considered as baseline. The peak-signal-to-noise ratio (PSNR) and structural similarity (SSIM) measures were averaged over ten sequences.

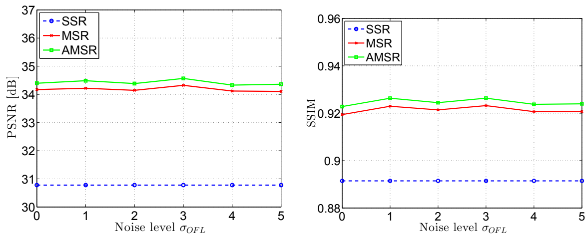


Figure 13: Robustness of multi-sensor (MSR) and adaptive multi-sensor super-resolution (AMSR) with respect to noise in photometric displacement fields. Single-sensor super-resolution (SSR) is considered as baseline. The peak-signal-to-noise ratio (PSNR) and structural similarity (SSIM) measures were averaged over ten sequences.

Noise in Displacement Fields. Next, we evaluated the robustness of the proposed MSR and AMSR methods with respect to noise in optical flow displacements fields. In practical applications, this situation appears in image regions without texture, where optical flow displacement fields are less reliable. For this experiment, we added zero-mean Gaussian noise to each component of the photometric displacement fields. The noise standard deviation was gradually increased from $\sigma_{\mathrm{OFL}} = 0$ to $\sigma_{\mathrm{OFL}} = 5.0$ in terms of pixels of the RGB images. The PSNR and SSIM measures were averaged for each level over ten sequences and plotted in Fig. 13. Even for severe noise levels, the proposed MSR and AMSR approaches reconstructed accurate super-resolved images in terms of the examined error measures.

6.2. Ex-Vivo Experiments

We examined our approach for image-guided surgery on real, ex-vivo data. First, we considered image guidance based on ToF imaging for open surgery. Second, we investigated super-resolution for 3-D endoscopy in minimally invasive surgery.

6.2.1. Quantitative Assessment

In order to assess the performance of super-resolution in the absence of ground truth range data, we employ two objective quality measures for quantitative evaluation. First, we evaluate the noise reduction on 3-D surfaces using a semi-automatic blind SNR estimation. This measure defined in dB is determined on flat surfaces selected manually in range data and computed according to:

$$Q_{\text{snr}} = 10 \log_{10} \left(\frac{\mu_{\text{flat}}}{\sigma_{\text{flat}}} \right), \tag{21}$$

where μ_{flat} and σ_{flat} denote the mean and standard deviation of the range values in the selected region of interest, respectively. High estimates Q_{snr} indicate a more accurate reconstruction of flat surfaces.

Moreover, we aim at accurate edge reconstruction as it is desired to reconstruct range data with sharp transitions between neighboring structures. In our semi-automatic approach, we manually select regions of interest with a transition between two structures. Range values in the selected region are modeled by a Gaussian mixture model (GMM) consisting of two components that represent foreground and background. Then, our edge reconstruction measure is computed by:

$$Q_{\text{edge}} = \frac{w_b(\mu_b - \mu)^2 + w_f(\mu_f - \mu)^2}{w_b\sigma_b^2 - w_f\sigma_f^2}, \tag{22}$$

where μ denotes the mean in the selected region, μ_b and μ_f are the mean range values of the background and the foreground, σ_b and σ_f are the corresponding standard deviations, and w_b and w_f denote the corresponding weights of the GMM. The GMM is fitted automatically using k-means clustering. Lower standard deviations σ_b and σ_f as well as higher distances between the mean values μ_b and μ_f indicate a better discrimination between the foreground and background range values, which leads to higher values of Q_{edge}.

6.2.2. Open Surgery

We measured a porcine liver using a PMD CamCube 3 with resolution of 200×200 px. Photometric information was obtained by a Grasshopper2 camera with a resolution of 1200×1200 px fixed on a tripod and combined with the ToF camera. For this stereo vision system, the method presented in Sect. 3.2 was used for fusion of both modalities. Subpixel motion during image acquisition was induced by small vibrations of the tripod. All parameters were the same as specified in Sect. 6.1.1 for synthetic data.

Super-resolution was employed for $K = 31$ frames and magnification $s = 4$. Qualitative results for a comparison of the different approaches are shown in Fig. 14. All super-resolution methods yielded an improved reconstruction of the liver surface compared to raw data degraded by a high amount of noise. However, in case of SSR edges are blurred as shown on the lobe of the liver. For MSR and AMSR, we observed similar results with an improved reconstruction of the liver boundary. The corresponding 3-D meshes rendered from raw data and the super-resolved data obtained by the AMSR method are presented in Fig. 16.

We carefully selected four image regions containing flat surfaces and four regions containing sharp edges in each dataset. In total, we utilized four datasets for quantitative evaluation. Using blind SNR estimates according to Eq. (21) averaged over all datasets, we observed substantial enhancements achieved by super-resolution compared to LR range data. In particular, AMSR achieved the best results for Q_{snr} given by 21.6 ± 1.08 dB. In terms of the edge measure defined in Eq. (22), the proposed multi-sensor methods achieved accurate edge reconstruction compared to SSR. Here, AMSR achieved the best results measured by $Q_{\text{edge}} = 6.75 \pm 1.55$. The statistics for both quantitative measures are summarized in Tab. 2.

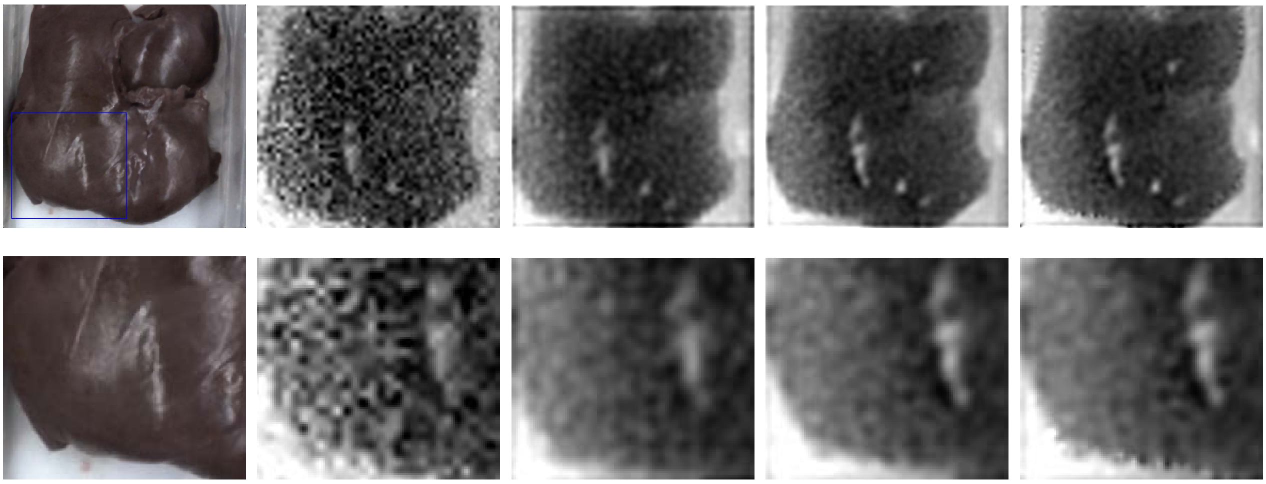
We evaluated the runtimes of super-resolution based on our non-parallelized Matlab implementation on an Intel Xeon E3-1245 CPU. For the aforementioned setup, the SSR method reconstructed one super-resolved image within $\approx$ 25 s including motion estimation and non-linear optimization. The MSR approach took $\approx$ 90 s due to the more demanding motion estimation on photometric data. The second optimization stage to implement the AMSR approach took $\approx$ 20 s.

6.2.3. Minimally Invasive Surgery

In terms of minimally invasive procedures, we examined hybrid 3-D endoscopy based on ToF imaging. In our ex-vivo experiments, we used a 3-D endoscope prototype manufactured by Richard Wolf GmbH, Knittlingen, Germany as used in (Köhler et al., 2013) for a phantom study. We acquired range data (64×48 px) and complementary RGB images (640×480 px) of a porcine liver. Endoscopic tools were included in each scene to simulate a realistic medical scenario. The homography approach

Table 2: Mean and standard deviation of the blind SNR measure Q_{snr} and the edge reconstruction measure Q_{edge} evaluated for the ToF/RGB endoscope as well as the PMD CamCube used in our open surgery setup. Both measures were evaluated for LR range data, single-sensor (SSR), multi-sensor (MSR) and adaptive multi-sensor super-resolution (AMSR).

Measure	**PMD CamCube**				**ToF/RGB endoscope**			
	LR data	**SSR**	**MSR**	**AMSR**	**LR data**	**SSR**	**MSR**	**AMSR**
Q_{snr}	17.9 ± 0.86	21.2 ± 1.21	21.5 ± 1.11	$\mathbf{21.6 \pm 1.08}$	7.20 ± 2.76	10.6 ± 1.94	10.7 ± 2.30	$\mathbf{10.9 \pm 1.99}$
Q_{edge}	3.85 ± 0.85	4.72 ± 1.02	6.71 ± 1.48	$\mathbf{6.75 \pm 1.55}$	1.89 ± 0.37	2.49 ± 0.76	2.87 ± 1.32	$\mathbf{2.96 \pm 1.66}$



(a) RGB image (b) LR range (c) SSR (d) MSR (e) AMSR

Figure 14: Results obtained from the surface of a porcine liver measured with a PMD CamCube 3 and a Grasshopper2 camera in a stereo setup for open surgery (magnification $s = 4$, $K = 31$ frames): RGB data (a) fused with LR range data (b) and the results obtained from the SSR (c), the MSR (d) and the AMSR approach (e) are shown in the top row. See the bottom row for a visual comparison of the evaluated methods in terms of noise reduction and edge reconstruction in a region of interest.

presented in Sect. 3.1 was utilized for sensor fusion. The endoscope was shifted and tools were slightly moved during acquisition to simulate movements of the surgeon. We used the same parameter setting as for synthetic data described in Sect. 6.1.1.

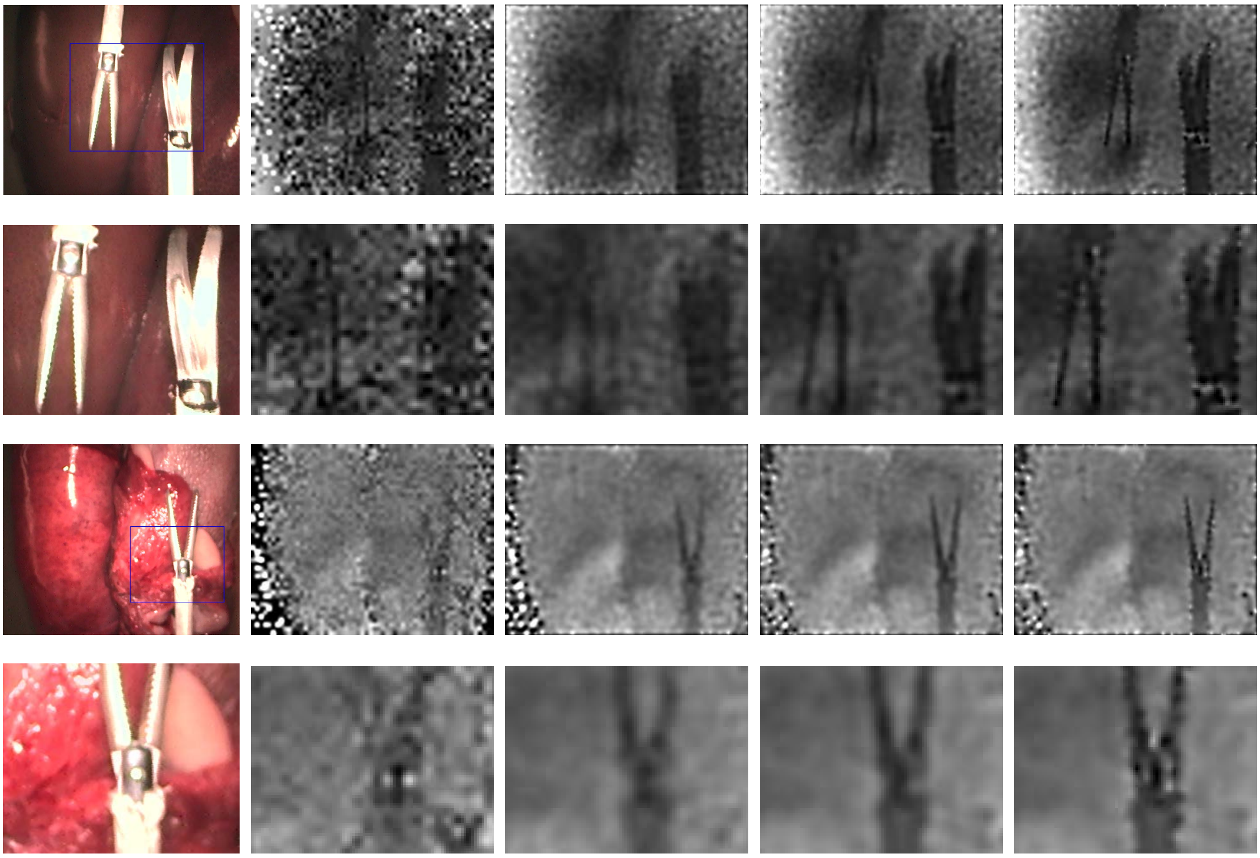
Super-resolution was performed for $K = 31$ frames and magnification $s = 4$. A comparison of the different methods is given in Fig. 15. Please see the improved reconstruction of endoscopic tools in the outcome of MSR barely visible in raw data and not accurately recovered by the SSR method. Additionally, in case of AMSR we observed enhanced edge reconstruction for these objects. For the associated 3-D meshes please see Fig. 16.

Similarly to our open surgery setup, we carefully selected four image regions containing flat surfaces and four regions containing sharp edges per dataset using nine ex-vivo datasets. The quantitative measures to assess noise reduction and edge reconstruction are reported in Tab. 2. As in our open surgery setup, super-resolution achieved a substantial enhancement of Q_{snr} compared to LR range data, whereas the different super-resolution approaches achieved similar results. Moreover, the proposed multi-sensor approaches, enhanced the edge reconstruction as quantified by Q_{edge}. The most accurate edge reconstruction was achieved by the proposed AMSR algorithm, where we obtained $Q_{\text{edge}} = 2.96 \pm 1.66$.

The runtime for SSR to obtain one super-resolved image was ≈ 20 s. For the same setup, the MSR approach took ≈ 220 s and the second optimization stage to of the AMSR approach took ≈ 40 s.

7. Discussion

We evaluated our proposed framework without (MSR) and with adaptive regularization (AMSR) for synthetic images as well as ex-vivo data.



(a) RGB (b) LR range (c) SSR (d) MSR (e) AMSR

Figure 15: Results obtained from a porcine liver (first row) and lung (third row) acquired with a 3-D ToF/RGB endoscope in our ex-vivo experiments (magnification $s = 4$, $K = 31$ frames): RGB data, LR range data and the outcomes of single-sensor (SSR), multi-sensor (MSR) as well as adaptive multi-sensor super-resolution (AMSR) are shown in the first and third row, respectively. See the second and fourth row for visual comparison in a region of interest.

7.1. Comparison to Single-Sensor Approach

Our experimental evaluation compared our multi-sensor framwork to super-resolution that exploits only range data. First, quantitative results summarized in Fig. 8 indicate increased accuracy of the proposed MSR approach compared to SSR. Optical flow estimation on photometric data provides accurate displacement fields for super-resolution, whereas SSR is susceptible to misregistration. In particular, improvements by our method are notably for higher magnification ($s = 4$), where the accuracy of motion estimation is a major limitation of super-resolution performance. In addition to accurate motion estimation, adaptive regularization enhances edge reconstruction compared to regularization with uniform weights as shown in Fig. 5. The benefit of this extension is demonstrated by improved PSNR and SSIM measures summarized in Fig. 8. In particular, improvements compared to the MSR approach are notably for higher magnification factors ($s = 4$). However, since this part of our framework enhances edges and does not affect flat areas, quantitative improvements in terms of PSNR and SSIM are minor compared to the differences between SSR and MSR. Due to a limited sample size and correlations between the observations, differences between MSR and AMSR are not statistically significant.

The proposed range correction scheme employed in our framework is essential to obtain unbiased range data in presence of general 3-D camera motion. As shown in Fig. 9, range correction becomes crucial for sequences with higher amount of out-of-plane motion (sequences S5 - S12), while there is no benefit in the absence of out-of-plane motion.

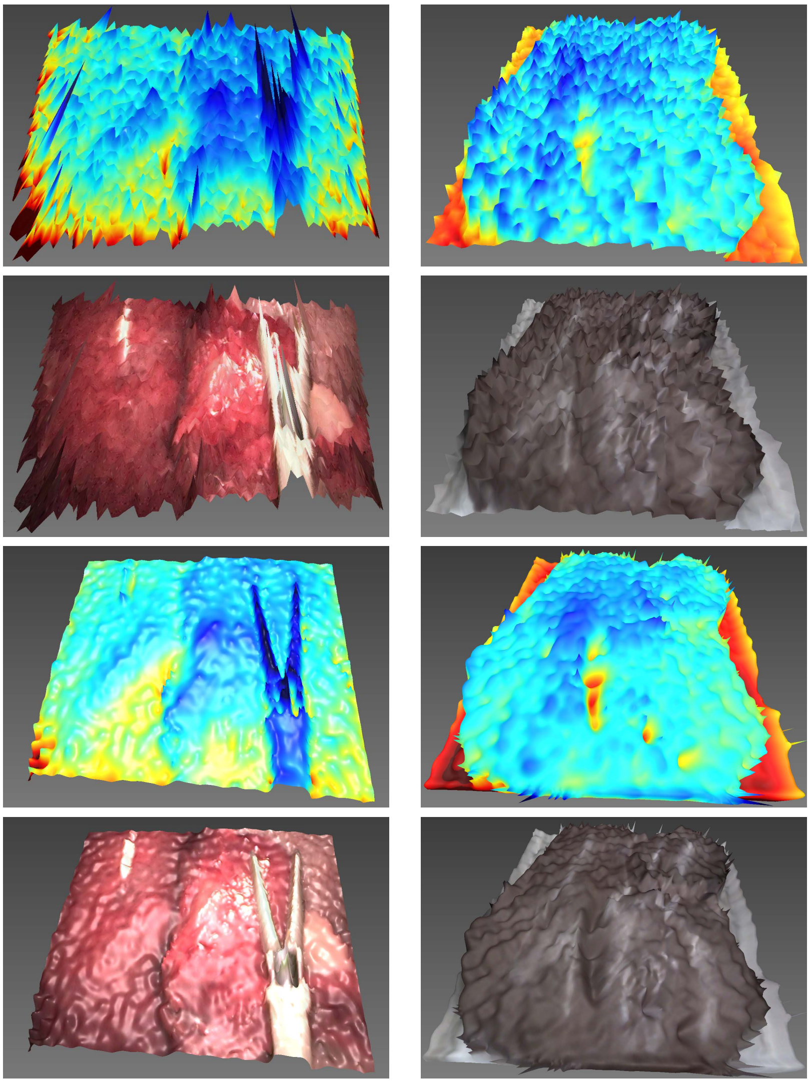


Figure 16: 3-D meshes rendered for the 3-D endoscope (left column) and the PMD CamCube (right column). First and third row: color coded texture for LR data and super-resolved data obtained from AMSR. Blue denoting values close to the sensor and red denoting values further away. Second and fourth row: photometric information used as overlay.

7.2. Robustness of our Framework

We also demonstrated the robustness of our framework under challenging and more realistic situations for image-guided surgery including degraded quality of photometric data as well as the uncertainty of sensor fusion and motion estimation.

Even if our framework relies on high-resolution photometric data, the proposed motion estimation scheme is still accurate for decreased spatial resolutions. As shown in Fig. 10, the proposed MSR approach outperforms SSR. The AMSR approach is more sensitive due to decreased robustness of edge detection on photometric data of low spatial resolution. Nonetheless, this is only severe for low spatial resolutions (Scale < 0.6, $384{\times}288$ px) that are considerably lower than in common hybrid RI setups as investigated in our work.

Results shown in Fig. 11 also demonstrate robustness of our framework in presence of specular reflections, which is a common issue in image-guided surgery. In case of more severe disturbances, e. g. smoke or highly dynamic scenes with non-rigid deformations, our approach may be susceptible to misregistration in individual frames. However, an appropriate outlier detection (Zhao and Sawhney, 2002; Köhler et al., 2014) may extend the proposed motion estimation scheme. Please note that our method is also able to recover from failure due to the underlying sliding window processing.

In terms of sensor fusion uncertainty, MSR is robust with respect to misalignments between range and photometric data as shown in Fig. 12. Similarly to related guided upsampling techniques (Kopf et al., 2007; He et al., 2010), AMSR is more sensitive since discontinuities in both modalities are no longer aligned. However, this only appears for severe misalignments that are avoided by accurate system calibration (Haase et al., 2013a).

An explicit evaluation of the uncertainty of estimated photometric displacements demonstrates the robustness of our framework in terms of motion estimation. As shown in Fig. 13, MSR and AMSR outperforms the SSR approach even for displacements fields affected by severe errors. This is essential for practical applications, where optical flow displacements in texture-less regions are unreliable.

7.3. Application to Image-Guided Surgery

We investigated our framework for minimally invasive and open surgery on ex-vivo data. Here, one additional issue is the presence of systematic, hardware dependent errors (Kolb et al., 2010) affecting LR range data. For SSR working solely on range data, this means an additional handicap especially for motion estimation. Our framework is more robust since motion estimation performed on high-quality photometric data is insensitive to these errors. In terms of 3-D endoscopy, this is visible by the improved reconstruction of endoscopic tools as shown in Fig. 15, which are barely visible in raw data and not accurately reconstructed by SSR. In particular, the most accurate reconstruction of edges in range data was obtained by AMSR utilizing adaptive regularization. For our stereo vision setup, we utilize a range sensor of higher spatial resolution. The proposed MSR and AMSR approaches substantially improve surface reconstruction as demonstrated in Fig. 14, while SSR suffers from blurred edges due to misregistrations. In this setup, the benefit gained by AMSR compared to MSR is smaller since surfaces are typically more smooth and thus edges cannot be exploited by adaptive regularization. Nevertheless, benefits in terms of edge reconstruction compared to SSR are noticeable.

The blind SNR estimate Q_{snr} provides a quantitative measure to assess noise reduction compared to LR range data. The improved Q_{snr} measure indicates an substantially enhanced surface reconstruction by means of super-resolution. This appears for 3-D endoscopy as well as for open surgery system, while in both setups the proposed AMSR method achieves the best results. Similarly, edge reconstruction can be quantitatively assessed by Q_{edge}. This measure demonstrates that our multi-sensor framework outperforms SSR, which is affected by blurring caused by inaccurate motion estimation. Considering all datasets used in our experiments, the most accurate edge reconstruction is achieved by the AMSR approach due to the underlying adaptive regularization scheme. This is an essential property to facilitate an automatic segmentation or registration based on super-resolved range data.

In terms of runtime, our experimental implementation does not enable real-time super-resolution limiting its application. However, recently it has been shown that the core of this framework including non-linear optimization can be considerably accelerated using the graphics processing unit (GPU) to enable real-time applications (Wetzl et al., 2013).

8. Conclusion and Future Work

This paper presents a multi-sensor super-resolution framework for hybrid imaging to super-resolve LR data by exploiting additional guidance images. In our work, this concept is applied in RI to enhance range images guided by complementary photometric data. We utilize sensor fusion to tackle motion estimation and regularization as essential issues towards robust super-resolution. The proposed method is applicable for hybrid 3-D endoscopy and image guidance in open surgery to overcome the poor resolution of range sensors as major limitation for clinical applications. Our multi-sensor approach achieves substantial image quality improvement in terms of spatial resolution as well as SNR and outperforms conventional single-sensor methods.

Future work will focus on the evaluation of our method for various image analysis tasks in image-guided surgery. In order to facilitate real-time applications, future work has to consider an acceleration of our approach, e. g. by means of GPU processing. Another issue often ignored in related work on range super-resolution is the modeling of systematic, sensor-specific errors. Here, the integration of correction schemes proposed e. g. by Lindner et al. (2010) or Reynolds et al. (2011) might further improve our super-resolution framework.

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. We gratefully acknowledge the support by the DFG under Grant No. HO 1791. Furthermore, this research was funded/supported by the Graduate School of Information Science in Health (GSISH), the TUM Graduate School and the research training group 1126 funded by the DFG. We thank the Metrilus GmbH for their support.

References

Baker, S., Kanade, T., 2002. Limits on super-resolution and how to break them. IEEE Transactions on Pattern Analysis and Machine Intelligence 24, 1167–1183.

Bauer, S., Seitel, A., Hofmann, H., Blum, T., Wasza, J., Balda, M., Meinzer, H.P., Navab, N., Hornegger, J., Maier-Hein, L., 2013. Real-Time Range Imaging in Health Care: A Survey, in: Time-of-Flight and Depth Imaging. Sensors, Algorithms, and Applications. Springer, pp. 228–254.

Bhavsar, A.V., Rajagopalan, A.N., 2012. Range map superresolution-inpainting, and reconstruction from sparse data. Computer Vision and Image Understanding 116, 572–591.

Bouguet, J.Y., 2013. Camera Calibration Toolbox for MATLAB. URL: `http://www.vision.caltech.edu/bouguetj/calib_doc/`.

Capel, D., 2004. Image mosaicing and super-resolution. Ph.D. thesis. Oxford University.

Cash, D.M., Miga, M.I., Glasgow, S.C., Dawant, B.M., Clements, L.W., Cao, Z., Galloway, R.L., Chapman, W.C., 2007. Concepts and preliminary data toward the realization of image-guided liver surgery. Journal of gastrointestinal surgery official journal of the Society for Surgery of the Alimentary Tract 11, 844–859.

Elad, M., Feuer, A., 1997. Restoration of a Single Superresolution Image from Several Blurred, Noisy, and Undersampled Measured Images. IEEE Transactions on Image Processing 6, 1646–1658.

Farsiu, S., Robinson, M., Elad, M., Milanfar, P., 2004. Fast and Robust Multiframe Super Resolution. IEEE Transactions on Image Processing 13, 1327–1344.

Field, M., Clarke, D., Strup, S., Seales, W., 2009. Stereo endoscopy as a 3-d measurement tool, in: EMBC 2009, pp. 5748–5751.

Fransens, R., Strecha, C., Van Gool, L., 2007. Optical flow based super-resolution: A probabilistic approach. Computer Vision and Image Understanding 106, 106–115.

Greenspan, H., 2008. Super-Resolution in Medical Imaging. The Computer Journal 52, 43–63.

Guo, W., Huang, F., 2008. A Local Mutual Information Guided Denoising Technique and Its Application to Self-calibrated Partially Parallel Imaging, in: Medical Image Computing and Computer-Assisted Intervention (MICCAI 2008), Springer-Verlag. pp. 939–947.

Haase, S., Forman, C., Kilgus, T., Bammer, R., Maier-Hein, L., Hornegger, J., 2013a. ToF/RGB Sensor Fusion for 3-D Endoscopy. Current Medical Imaging Reviews 9, 113–119.

Haase, S., Wasza, J., Kilgus, T., Hornegger, J., 2013b. Laparoscopic instrument localization using a 3-d time-of-flight/rgb endoscope, in: WACV 2013, pp. 449–454.

Hardie, R., Barnard, K., Armstrong, E., 1997. Joint MAP registration and high-resolution image estimation using a sequence of undersampled images. IEEE Transactions on Image Processing 6, 1621–1633.

He, K., Sun, J., Tang, X., 2010. Guided Image Filtering, in: Proceedings of European conference on computer vision: Part I, pp. 1–14.

Irani, M., Peleg, S., 1991. Improving Resolution by Image Registration. CVGIP: Graphical Models and Image Processing 53, 231–239.

Kennedy, J.A., Israel, O., Frenkel, A., Bar-Shalom, R., Azhari, H., 2007. Improved image fusion in PET/CT using hybrid image reconstruction and super-resolution. International Journal of Biomedical Imaging 2007.

Kilgus, T., Heim, E., Haase, S., Prfer, S., Mller, M., Seitel, A., Fangerau, M., Wiebe, T., Iszatt, J., Schlemmer, H.P., Hornegger, J., Yen, K., Maier-Hein, L., 2014. Mobile markerless augmented reality and its application in forensic medicine. International Journal of Computer Assisted Radiology and Surgery , 1–14.

Köhler, T., Haase, S., Bauer, S., Wasza, J., Kilgus, T., Maier-Hein, L., Feußner, H., Hornegger, J., 2013. Tof meets rgb: Novel multi-sensor super-resolution for hybrid 3-d endoscopy, in: Medical Image Computing and Computer-Assisted Intervention (MICCAI 2013). Springer Berlin Heidelberg. volume 8149, pp. 139–146.

Köhler, T., Haase, S., Bauer, S., Wasza, J., Kilgus, T., Maier-Hein, L., Feußner, H., Hornegger, J., 2014. Outlier detection for multi-sensor super-resolution in hybrid 3d endoscopy, in: Bildverarbeitung fr die Medizin (BVM 2014). Springer Berlin Heidelberg, pp. 84–89.

Kolb, A., Barth, E., Koch, R., Larsen, R., 2010. Time-of-Flight Cameras in Computer Graphics. Computer Graphics Forum 29, 141–159.

Kopf, J., Cohen, M.F., Lischinski, D., Uyttendaele, M., 2007. Joint Bilateral Upsampling. ACM Transactions on Graphics 26.

Lin, Z., Shum, H.Y., 2004. Fundamental limits of reconstruction-based superresolution algorithms under local translation. IEEE Transactions on Pattern Analysis and Machine Intelligence 26, 83–97.

Lindner, M., Schiller, I., Kolb, A., Koch, R., 2010. Time-of-Flight sensor calibration for accurate range sensing. Computer Vision and Image Understanding 114, 1318–1328.

Liu, C., 2009. Beyond Pixels: Exploring New Representations and Applications for Motion Analysis. Ph.D. thesis. Massachusetts Institute of Technology.

Maier-Hein, L., Groch, A., Bartoli, A., Bodenstedt, S., Boissonnat, G., Chang, P.L., Clancy, N.T., Elson, D.S., Haase, S., Heim, E., Hornegger, J., Jannin, P., Kenngott, H., Kilgus, T., Müller-Stich, B., Oladokun, D., Röhl, S., Dos Santos, T.R., Schlemmer, H.P., Seitel, A., Speidel, S., Wagner, M., Stoyanov, D., 2014. Comparative validation of single-shot optical techniques for laparoscopic 3-D surface reconstruction. IEEE transactions on medical imaging 33, 1913–30.

Maier-Hein, L., Schmidt, M., Franz, A.M., dos Santos, T.R., Seitel, A., Jähne, B., Fitzpatrick, J.M., Meinzer, H.P., 2010. Accounting for anisotropic noise in fine registration of time-of-flight range data with high-resolution surface data, in: Medical Image Computing and Computer-Assisted Intervention (MICCAI 2010), pp. 251–258.

Mersmann, S., Müller, M., Seitel, A., et al., 2011. Time-of-flight camera technique for augmented reality in computer-assisted interventions, in: Proc SPIE, pp. 79642C–79642C.

Milanfar, P., 2010. Super-Resolution Imaging. CRC Press.

Mitzel, D., Pock, T., Schoenemann, T., Cremers, D., 2009. Video Super Resolution using Duality Based TV-L 1 Optical Flow. Science , 1–10.

Nabney, I.T., 2002. NETLAB: Algorithms for Pattern Recognition. Advances in Pattern Recognition. 1^{st} ed., Springer.

Nguyen, N., Milanfar, P., Golub, G., 2001. Efficient generalized cross-validation with applications to parametric image restoration and resolution enhancement. IEEE Transactions on Image Processing 10, 1299–308.

Park, J., Kim, H., Tai, Y.W., Brown, M., Kweon, I., 2011. High quality depth map upsampling for 3D-TOF cameras, in: IEEE International Conference on Computer Vision, pp. 1623–1630.

Penne, J., Höller, K., Stürmer, M., Schrauder, T., Schneider, A., Engelbrecht, R., Feußner, H., Schmauss, B., Hornegger, J., 2009. Time-of-Flight 3-D Endoscopy, in: Medical Image Computing and Computer-Assisted Intervention (MICCAI 2009), pp. 467–474.

Pickup, L.C., Capel, D.P., Roberts, S.J., Zisserman, A., 2007. Overcoming registration uncertainty in image super-resolution: maximize or marginalize? EURASIP Journal on Advances in Signal Processing 2007, 20–20.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., R Core Team, 2014. nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-118.

Pluim, J.P.W., Maintz, J.B.A., Viergever, M.A., 2003. Mutual-information-based registration of medical images: a survey. IEEE Transactions on Medical Imaging 22, 986–1004.

Reynolds, M., Dobos, J., Peel, L., Weyrich, T., Brostow, G., 2011. Capturing time-of-flight data with confidence, in: 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 945–952.

Robinson, D., Farsiu, S., Milanfar, P., 2009. Optimal Registration Of Aliased Images Using Variable Projection With Applications To Super-Resolution. The Computer Journal 52, 31–42.

Schmalz, C., Forster, F., Schick, A., Angelopoulou, E., 2012. An endoscopic 3D scanner based on structured light. Medical Image Analysis 16, 1063–1072.

Schuon, S., Theobalt, C., Davis, J., Thrun, S., 2009. LidarBoost: Depth superresolution for ToF 3D shape scanning, in: IEEE Conference on Computer Vision and Pattern Recognition, pp. 343–350.

Sun, H., Lunn, K.E., Farid, H., Wu, Z., Roberts, D.W., Hartov, A., Paulsen, K.D., 2005. Stereopsis-guided brain shift compensation. IEEE Transactions on Medical Imaging 24, 1039–52.

Tipping, M.E., Bishop, C.M., 2003. Bayesian Image Super-resolution, in: Advances in Neural Information Processing Systems, pp. 1303–1310.

Torr, P.H.S., Zisserman, A., 2000. MLESAC: A New Robust Estimator with Application to Estimating Image Geometry. Computer Vision and Image Understanding 78, 138–156.

Tsai, R., Huang, T., 1984. Multipleframe image restoration and registration. Advances in Computer Vision and Image Processing , 317–339.

Vandewalle, P., Süsstrunk, S., Vetterli, M., 2006. A frequency domain approach to registration of aliased images with application to super-resolution. EURASIP Journal on Applied Signal Processing 2006, 233–233.

Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P., 2004. Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing 13, 600–612.

Wasza, J., Bauer, S., Haase, S., Schmid, M., Reichert, S., Hornegger, J., 2011a. RITK: The Range Imaging Toolkit - A Framework for 3-D Range Image Stream Processing, in: Proceeding of International Workshop on Vision, Modeling and Visualization (VMV), pp. 57–64.

Wasza, J., Bauer, S., Hornegger, J., 2011b. Real-time preprocessing for dense 3-D range imaging on the GPU: Defect interpolation, bilateral temporal averaging and guided filtering, in: IEEE International Conference on Computer Vision Workshops, pp. 1221–1227.

Wetzl, J., Taubmann, O., Haase, S., Köhler, T., Kraus, M., Hornegger, J., 2013. Gpu-accelerated time-of-flight super-resolution for image-guided surgery, in: Bildverarbeitung für die Medizin (BVM 2013). Springer Berlin Heidelberg, pp. 21–26.

Yuan, Q., Zhang, L., Shen, H., 2012. Multiframe Super-Resolution Employing a Spatially Weighted Total Variation Model. IEEE Transactions on Circuits and Systems for Video Technology 22, 379–392.

Zhao, W., Sawhney, H.S., 2002. Is Super-Resolution with Optical Flow Feasible?, in: Computer Vision ECCV 2002. Springer Berlin Heidelberg. volume 2350 of *Lecture Notes in Computer Science*, pp. 599–613.

Zomet, A., Peleg, S., 2002. Multi-sensor super-resolution, in: Sixth IEEE Workshop on Applications of Computer Vision, 2002 (WACV 2002), pp. 27–31.