

GPU-Accelerated Time-of-Flight Super-Resolution for Image-Guided Surgery

Jens Wetzl^{1,*}, Oliver Taubmann^{1,*}, Sven Haase¹, Thomas Köhler^{1,2},
Martin Kraus^{1,2}, Joachim Hornegger^{1,2}

¹Pattern Recognition Lab, FAU Erlangen-Nuremberg

²Erlangen Graduate School in Advanced Optical Technologies (SAOT)

*These authors contributed equally to this work

jens.wetzl@fau.de

Abstract. In the field of image-guided surgery, Time-of-Flight (ToF) sensors are of interest due to their fast acquisition of 3-D surfaces. However, the poor signal-to-noise ratio and low spatial resolution of today's ToF sensors require preprocessing of the acquired range data. Super-resolution is a technique for image restoration and resolution enhancement by utilizing information from successive raw frames of an image sequence. We propose a super-resolution framework using the graphics processing unit. Our framework enables interactive frame rates, computing an upsampled image from 10 noisy frames of 200×200 px with an upsampling factor of 2 in 109 ms. The root-mean-square error of the super-resolved surface with respect to ground truth data is improved by more than 20% relative to a single raw frame.

1 Introduction

Image-guided surgery provides physicians with helpful information and thus speeds up and improves medical interventions. One pertinent example is enhanced surface representations for augmented reality applications [1]. Time-of-Flight (ToF) sensors hold great potential for acquiring 3-D surfaces during an intervention due to their fast and dense acquisition technique. However, these sensors still suffer from low spatial resolution compared to state-of-the-art color sensors. Furthermore, high temporal and spatial noise in the range data is a major issue. To compensate for this, various preprocessing and calibration techniques have been proposed [2, 3]. Besides these approaches, super-resolution techniques present a promising alternative with the capability to improve noisy range data while increasing their spatial resolution. The goal of multi-frame super-resolution is to fuse several low-resolution (LR) frames into one high-resolution (HR) image while preserving edges and suppressing noise [4]. Each LR frame shows the scene from a slightly different viewpoint. The motion between successive frames is utilized to obtain a finer sampling compared to a single image.

As shown by Schuon et al. [5], the quality of range images can be improved significantly by super-resolution. However, their approach uses a simplified imaging

model where the point spread function (PSF) to model sampling of a camera is not taken into account and only translational motion between successive frames is assumed. The application of this method to image-guided surgery at interactive frame rates is infeasible due to prohibitive computational effort. General purpose super-resolution in real time has been demonstrated in [6], proposing an interpolation-based scheme with no appropriate physical model for image generation and a restriction to translational motion.

In this paper, we present a framework capable of recovering HR images from a series of preregistered LR images at interactive frame rates for intraoperative image restoration. Our method is based on a generative image model and formulated as a nonlinear optimization problem. The imaging model used for super-resolution covers affine motion and a Gaussian PSF. All steps of the algorithm are accelerated using the graphics processing unit (GPU) with Nvidia’s CUDA platform to enable image-guided surgery at interactive frame rates.

2 Materials and methods

The super-resolution framework presented in this paper is based on a maximum a posteriori (MAP) estimate of the desired HR image. We use a limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) optimizer for minimization of the MAP cost function (Fig. 1).

2.1 Maximum a posteriori super-resolution

Our method is based on a forward model that describes the generation of LR frames from the ideal HR image that should be recovered. Let $\mathbf{x} \in \mathbb{R}^N$ be an HR image where the pixels are arranged in linear order. The k^{th} LR frame $\mathbf{y}^{(k)} \in \mathbb{R}^M$, $M < N$, out of a sequence of K frames $\mathbf{y}^{(1)}, \dots, \mathbf{y}^{(K)}$ is related to \mathbf{x} according to

$$\mathbf{y}^{(k)} = \mathbf{W}^{(k)}\mathbf{x} + \boldsymbol{\varepsilon}^{(k)} \quad (1)$$

where $\mathbf{W}^{(k)}$ denotes the system matrix which models warping, blur and decimation of the HR image and $\boldsymbol{\varepsilon}^{(k)}$ is zero-mean Gaussian noise corrupting the k^{th} LR

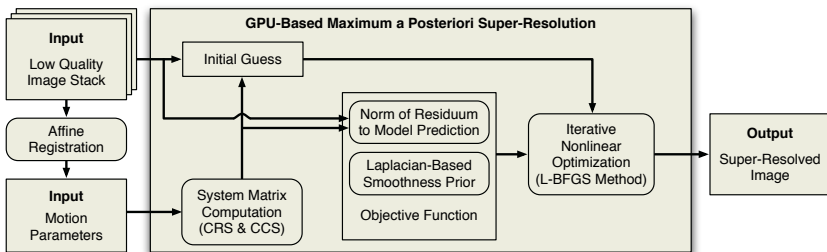


Fig. 1. System overview of our super-resolution framework.

image. Each LR frame $\mathbf{y}^{(k)}$ for $k \geq 2$ is related to $\mathbf{y}^{(1)}$ by an affine homography $\mathbf{H}^{(k)} \in \mathbb{R}^{3 \times 3}$ such that $\mathbf{u}' = \mathbf{H}^{(k)} \mathbf{u}$, where \mathbf{u} and \mathbf{u}' are homogeneous pixel coordinates in $\mathbf{y}^{(k)}$ and $\mathbf{y}^{(1)}$ respectively. Modeling blur caused by the camera with a Gaussian PSF, the elements of the system matrix are given by

$$W_{mn} = \exp \left(- \frac{(\mathbf{v}_n - \mathbf{u}'_m)^T \nabla \mathbf{H} (\nabla \mathbf{H})^T (\mathbf{v}_n - \mathbf{u}'_m)}{2\sigma^2} \right) \quad (2)$$

where σ specifies the width of the PSF, $\nabla \mathbf{H}$ denotes the Jacobian of the affine transform $\mathbf{H}\mathbf{u}$ with respect to $\mathbf{u} = (u_1, u_2, 1)$, \mathbf{v}_n is the position of the n^{th} HR pixel and \mathbf{u}'_m is the position of the m^{th} LR pixel transformed into the HR coordinate system [4]. The rows of the system matrix are normalized to unity.

The super-resolved image \mathbf{x}^* is the minimum of the objective function

$$\mathbf{x}^* = \operatorname{argmin}_{\mathbf{x}} \left(\|\mathbf{W}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda \cdot \|\mathbf{h}_\delta(\mathbf{D}\mathbf{x})\|_1 \right) \quad (3)$$

where \mathbf{y} and \mathbf{W} are the stacked LR images and system matrices respectively. For regularization, we use the pseudo-Huber loss function $\mathbf{h}_\delta(\cdot)$ applied element-wise as $h_\delta(a) = \delta^2(\sqrt{1 + (a/\delta)^2} - 1)$ on the Laplacian $\mathbf{D}\mathbf{x}$ of the HR image \mathbf{x} , with λ controlling the strength of the prior. This imposes smoothness on the super-resolved image and prevents our estimation from converging to undesirable solutions where noise is amplified. The minimum of the objective function is a MAP estimate for the desired HR image [4]. Minimization is performed by the Quasi-Newton optimizer described in section 2.2.

We note that all system matrix elements can be computed independently according to (2). For better load-balancing, one GPU thread per matrix row is used. To reduce the memory footprint, W_{mn} is set to zero beyond 3 standard deviations of the point spread function, enabling us to store \mathbf{W} efficiently in a sparse matrix format. Furthermore, the Laplacian filtered image and the pseudo-Huber prior are calculated pixel-wise in parallel.

2.2 L-BFGS optimizer

As the gradient of the objective function given in (3) is nonlinear in the pixels of the HR image, we use an iterative L-BFGS optimizer [7] for nonlinear minimization. This algorithm is one of the most popular members of the family of Quasi-Newton methods and does not store a dense approximation of the Hessian, making it well-suited for a GPU implementation with limited memory. Despite relying on a low-rank representation, it is known to converge very quickly, which we could also confirm in our experiments. Since the overall objective function is convex, the minimum obtained by L-BFGS is an optimal solution. As initial guess for minimization, we choose an ‘‘average image’’ of the registered LR images, computed as $\widetilde{\mathbf{W}}^T \mathbf{y}$ where $\widetilde{\mathbf{W}}$ is \mathbf{W} with normalized columns [4].

2.3 Experiments

The proposed framework was implemented based on Nvidia’s CUDA platform and is available on our website¹. All vectorial computations are performed on the GPU, obviating the need to frequently transfer large amounts of data between host and device memory. The cuBLAS library is used for standard operations like dot products, vector scaling and scaled multiply-add. The system matrix \mathbf{W} is stored in both compressed row storage (CRS) and compressed column storage (CCS) formats as it is necessary to compute matrix-vector products for both \mathbf{W} and its transpose efficiently. Sparse matrix multiplication and the CRS/CCS management are handled by the cuSPARSE library.

In our ex-vivo experiments, we measured a porcine liver with a PMD Cam-Cube 3.0 and a high-precision light-section sensor [8] simultaneously. The data acquired by the latter served as ground truth.

For quantitative evaluation, absolute distance statistics and the root-mean-square error (RMSE) between the super-resolution output and the ground truth data were calculated. For this purpose, we registered both using optical markers. To obtain robust results, 240 ToF frames were split into 15 even sets. For each set, super-resolution was performed with different upsampling factors and sequence lengths. Both runtimes and error measurements were then averaged over all sets. All calculations were performed on an Nvidia GTX 580. The required image registration was performed offline using a 2D affine registration framework [9].

3 Results

In Fig. 2, we compare nearest-neighbor upsampling of an LR frame to a pre-processed [2] and our super-resolved image. Note that our framework has the desirable property of preserving edges as compared to the state-of-the-art pre-processing system, which becomes even more evident with larger motion. Fig. 3 shows a 3D mesh reconstruction of both the raw data and our result.

¹ <http://www5.cs.fau.de/research/software/>

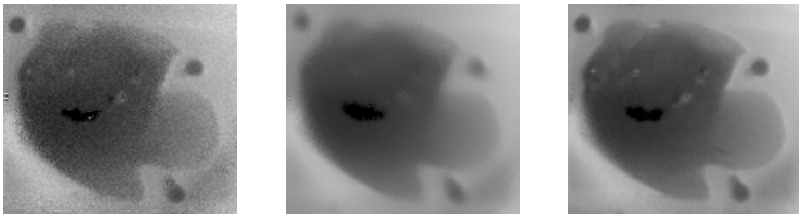


Fig. 2. Nearest neighbor (left) and preprocessed (middle) upsampling of one low-resolution range image, the range image obtained by super-resolution using 10 frames and upsampling factor of 2 (right).

	Raw	SR
RMSE	8.11 mm	6.29 mm
Median	4.92 mm	3.16 mm
Std.dev.	± 0.34 mm	± 0.16 mm

Table 1. Root-mean-square error (RMSE) and median as well as standard deviation of absolute distances to the ground truth mesh, averaged over 15 sets. Parameters for SR: 10 input frames and $2\times$ upsampling factor.

Quantitative results regarding the errors are given in Tab. 1. All measured statistical properties were improved by our super-resolution framework. Further evaluation is performed on runtimes for different sequence lengths and upsampling factors (Fig. 5).

4 Discussion

In this paper, we presented a framework for super-resolution of preregistered ToF range images at interactive frame rates, running on off-the-shelf GPUs. Experiments on porcine liver data acquired with a PMD CamCube showed promising results regarding both accuracy and performance. We were able to decrease the RMSE by more than 20% on average from a single raw frame by upsampling 10



Fig. 3. The porcine liver measured for all experiments (left) and reconstructed meshes from raw data (middle) and after super-resolution of 10 frames with $2\times$ zoom factor (right).

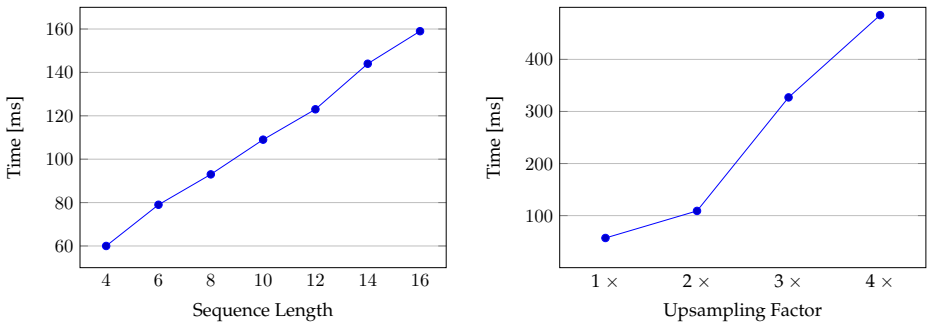


Fig. 4. Runtime analysis of our framework dependent on the number of input frames for an upsampling factor of 2 (left) and on the upsampling factor with 10 frames (right).

frames of 200×200 px. We computed the super-resolved image with an upsampling factor of 2 in 109 ms. Interactive frame rates were achieved for all evaluated sequence lengths given an upsampling factor of 2.

Future work will have to focus on integrating robust affine registration for a complete GPU accelerated system. Prior work by Ansorge et al. [10], using affine registration as an initialization for B-Spline registration in CUDA, strongly suggests that this is attainable. More sophisticated models that combine the registration step with the super-resolution optimization process have been proposed and may also be considered for use within our system.

Acknowledgement. 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, as well as the support by the DFG under Grant No. HO 1791/7-1. This research was funded/supported by the Graduate School of Information Science in Health (GSISH) and the TUM Graduate School.

References

1. Cash D, Miga M, Glasgow S, et al. Concepts and preliminary data toward the realization of image-guided liver surgery. *J Gastrointest Surg.* 2007;11(7):844–59.
2. Wasza J, Bauer S, Hornegger J. Real-time preprocessing for dense 3-D range imaging on the GPU: defect interpolation, bilateral temporal averaging and guided filtering. *Proc IEEE Int Conf Comput Vis.* 2011; p. 1221–7.
3. Fuchs S, Hirzinger G. Extrinsic and depth calibration of ToF-cameras. *Proc IEEE CVPR.* 2008; p. 1–6.
4. Pickup LC. Machine learning in multi-frame image super-resolution; 2008. PhD Thesis, University of Oxford.
5. Schuon S, Theobalt C, Davis J, et al. High-quality scanning using time-of-flight depth superresolution. *Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit.* 2008; p. 1–7.
6. Patil VH, Bormane DS, Patil HK. Real time super resolution image reconstruction. *Proc ICIAS.* 2007; p. 651–4.
7. Liu DC, Nocedal J. On the limited memory BFGS method for large scale optimization. *Math Program.* 1989;45(3):503–28.
8. Ettl S, Arold O, Yang Z, et al. Flying triangulation—an optical 3D sensor for the motion-robust acquisition of complex objects. *Appl Opt.* 2012;51(2):281–9.
9. Evangelidis GD, Psarakis EZ. Parametric image alignment using enhanced correlation coefficient maximization. *IEEE Trans Pattern Anal Mach Intell.* 2008;30:1858–65.
10. Ansorge RE, Sawiak SJ, Williams GB. Exceptionally fast non-linear 3D image registration using GPUs. *IEEE Nucl Sci Symp Conf Rec.* 2009; p. 4088–94.