

# LidarBoost: Depth Superresolution for ToF 3D Shape Scanning

Sebastian Schuon  
Stanford University

[schuon@cs.stanford.edu](mailto:schuon@cs.stanford.edu)

Christian Theobalt  
Stanford University

[theobalt@cs.stanford.edu](mailto:theobalt@cs.stanford.edu)

James Davis  
UC Santa Cruz

[davis@cs.ucsc.edu](mailto:davis@cs.ucsc.edu)

Sebastian Thrun  
Stanford University

[thrun@stanford.edu](mailto:thrun@stanford.edu)

## Abstract

*Depth maps captured with time-of-flight cameras have very low data quality: the image resolution is rather limited and the level of random noise contained in the depth maps is very high. Therefore, such flash lidars cannot be used out of the box for high-quality 3D object scanning. To solve this problem, we present LidarBoost, a 3D depth superresolution method that combines several low resolution noisy depth images of a static scene from slightly displaced viewpoints, and merges them into a high-resolution depth image. We have developed an optimization framework that uses a data fidelity term and a geometry prior term that is tailored to the specific characteristics of flash lidars. We demonstrate both visually and quantitatively that LidarBoost produces better results than previous methods from the literature.*

## 1. Introduction

The ability to capture decent 3D models of static real world objects has reached increased importance in many fields of application, such as manufacturing and prototyping, but also in the design of virtual worlds for movies and games. To capture a complete model of a scene, many individual scans have to be taken from different viewpoints and finally merged into a complete shape representation. It is no wonder that the choice of sensor for such a task is highly important.

Recently, new time-of-flight (ToF) cameras or flash lidars have been introduced that capture 3D depth maps by measuring the return travel time of an infrared light wavefront emitted from the sensor. At a first glance, these cameras seem to be very suitable for 3D shape scanning, as they can capture hundreds of depth scans in only a few seconds. Unfortunately, the resolution of the depth maps is far too low and the level of random noise is of such significance that, out-of-the-box, flash lidars cannot be used for 3D object scanning.

To overcome this problem, we develop in this paper a new 3D depth sensor superresolution approach. The core idea is to take a handful of depth images of a static scene

from only slightly displaced viewpoints (such that parallax can be neglected), align them, and combine them into a single high-resolution depth image. The resulting 3D depth map is of sufficient quality for 3D shape scanning, since it has much higher resolution than any input image, and measurement noise has been canceled out. The main contributions of this paper are

- A 3D depth sensor superresolution method that incorporates ToF specific knowledge and data. Additionally a new 3D shape prior is proposed, that enforces 3D specific properties.
- A comprehensive evaluation of the working range and accuracy of our algorithm using synthetic and real data captured with a ToF camera.
- Only few depth superresolution approaches have been developed previously. We show that our algorithm clearly outperforms the most related approaches.

## 2. Related Work

Time-of-flight cameras, such as the MESA Swissranger™, enable full-frame capture of 3D depth maps of general scenes at video rate. Furthermore, measurement accuracy is largely independent of surface texture when compared to passive reconstruction methods like stereo [13].

Greater flexibility and fast capture rates of ToF sensors, however, come at the price of low X/Y sensor resolution and often significant random measurement noise. To get quality recordings of higher resolution from noisy depth recordings, superresolution algorithms can be applied. Only a few approaches dealing with superresolution of any 3D scanner have been presented so far, which roughly fall into two categories:

**Combining Depth and Image Data** One strategy to up-sample and denoise a 3D recording leverages information from a high-resolution image of the same scene that was taken from a viewpoint close to the depth sensor. The core idea is to enforce simple statistical relations between depth and intensity data, such as the joint occurrence of depth and intensity edges, and smoothness of geometry in areas

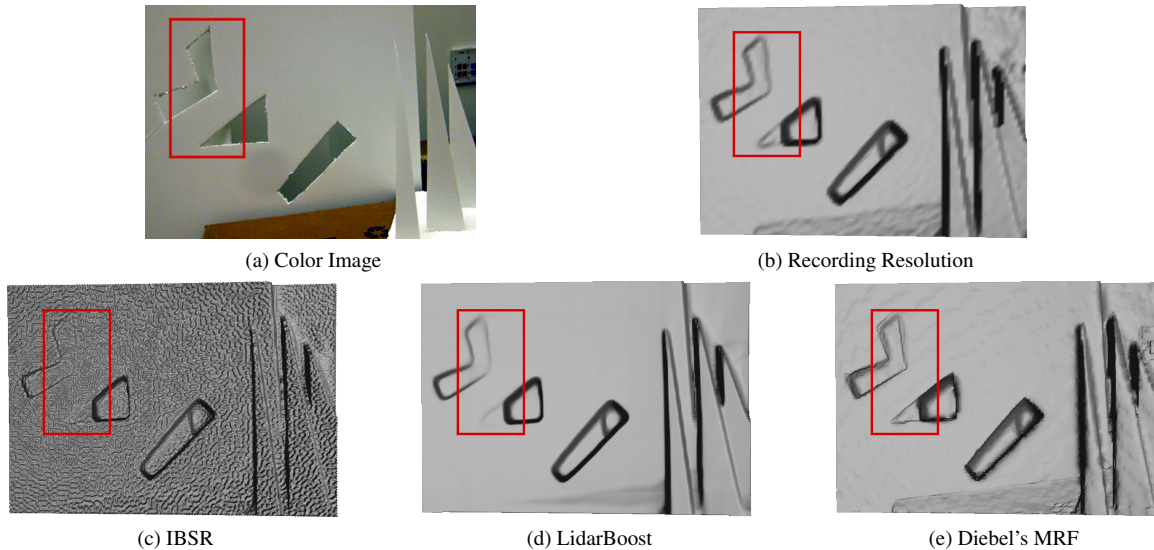


Figure 1: Real scene - wedges and panels (a): This scene with many depth edges (b) demonstrates the true resolution gain. IBSR (c) demonstrates increased resolution at the edges, but some aliasing remains and the strong pattern in the interior persists. LidarBoost (d) reconstructs the edges much more clearly and there is hardly a trace of aliasing, also the depth layers visible in the red encircled area are better captured. MRF upsampling (e) oversmooths the depth edges and in some places allows the low resolution aliasing to persist.

of largely uniform color [2, 14, 7]. Although these methods are often computationally very efficient, they frequently suffer from artifacts that are due to the heuristic nature of the enforced statistical model, e.g. copying of intensity texture into 3D geometry. Also, having to incorporate a second type of sensor may restrict the application range. Our method only requires depth images and does not suffer from texture copying.

**Superresolution From Depth Data Only** Here, the goal is to enhance the resolution by combining only depth recordings of a static scene that were taken from slightly displaced viewpoints. Kil et al. [5] were among the first to explore such an idea for laser triangulation scanners. They heavily oversample the scene by taking up to 100 scans from similar viewpoints to achieve four-times upsampled geometry. Since their data is so dense, and the random noise level of a laser scanner is significantly lower than that of a ToF camera, they can obtain good results by regular resampling from the aligned scan points with associated Gaussian location uncertainty. Reportedly, results may exhibit unnecessary blur and it is unlikely that this data fusion principle will work for highly noisy ToF data.

Only recently researchers looked into performing super-resolution on ToF camera data. Rajagopalan et al. [10] proposed a Markov-Random-Field based resolution enhancement method from a set of low-resolution depth recordings that formulates the upsampled 3D geometry as the most likely surface given several low resolution measurements. Their MRF uses a neighborhood system that en-

forces an edge-preserving smoothness prior between adjacent depth pixels. Their formulation of the problem bears two disadvantages: first complex parameter selection and secondly the formulation of the prior renders the problem non-convex, and hence more sophisticated solvers are required.

In another line of thinking, researchers tried to apply ideas from image superresolution algorithms, which combine low resolution intensity images into a high-resolution intensity image, to the 3D case. At a first glance, this seems like a reasonable strategy since for many years ever more successful image superresolution methods have been developed [9]. This strategy was first used by Rosenbush et al. [11]. While their paper was still very focused on alignment algorithms for depth measurements, they showed that ideas from image superresolution can in general be applied towards depth measurements. However, their reconstruction is rather basic and simply aligns the low resolution measurements on the high resolution grid and interpolates missing points.

Lately, more advanced image superresolution methods have been proposed that may also function in the 3D domain. Schuon et al. [12] verified this and applied the image superresolution method by Farsiu et al. [3] to depth images taken with a 3DV<sup>TM</sup> ToF camera. Following a general principle put forward by many image superresolution methods, they solve an energy minimization problem that jointly employs a *data term*, enforcing similarity between the input and output images, and a bilateral *regularization term* for

edge-preserving smoothness.

Our algorithm also falls into the latter category, but clearly outperforms previous approaches in terms of reconstruction quality and accuracy.

### 3. Our Algorithm

In our measurement setup we capture  $N$  depth images of a static scene,  $Y_k \in \mathbb{R}^{n \times m}$ , each having depth sensor resolution  $n \times m$ . Each depth image (also called depth map) is a grid of depth pixels, where each depth pixel records the distance to a 3D scene point along the measurement ray through the pixel. Given intrinsic ToF camera calibration data, a depth map can be reprojected into 3D geometry in world space. The  $Y_k$  are captured from only slightly displaced viewpoints which is why parallax effects can be neglected. Prior to superresolution, all depth images are registered against a reference frame out of  $Y_k$ . Reliable registration is important for superresolution, but in our case simple a procedure sufficed. Once registered, we compute a single high resolution depth image with  $\beta$  times higher resolution  $X \in \mathbb{R}^{\beta n \times \beta m}$  by solving an optimization problem of the form:

$$\text{minimize } E_{\text{data}}(X) + E_{\text{regular}}(X).$$

The first term  $E_{\text{data}}(X)$  is a data term measures agreement of the reconstruction with the aligned low resolution maps, Sect. 3.1.  $E_{\text{regular}}(X)$  is a regularization or prior energy term that guides the optimizer towards plausible 3D reconstructions if data points are sparse, Sect. 3.2.

This formulation is common to most superresolution methods. However their data and prior terms are designed for intensity images and cause strong artifacts when applied to depth images, as shown in Fig. 1c for the example of Schuon et al.’s method [12]. In contrast, our prior and data terms explicitly take into account the specifics of the 3D reconstruction problem as well as the characteristics of the time-of-flight sensors used. In contrast to related 3D up-sampling methods, our formulation yields a convex optimization problem which makes the superresolution procedure efficient and robust. Overall, our superresolved depth maps therefore exhibit a much higher quality than it was achieved with previous approaches for ToF superresolution.

#### 3.1. Data Term

The data term ensures that the final superresolved depth map is coherent with the registered low resolution measurements  $Y_k \in \mathbb{R}^{n \times m}$ . During preprocessing,  $N - 1$  frames out of the  $Y_k$  frames are aligned against a reference frame by computing for each a displacement vector. Typically, the first frame from  $Y_k$  is chosen as reference frame. Currently, we use hierarchical Lukas Kanade optical flow [8] to compute the registration but alternative registration ap-

proaches would be feasible. This process and the upsampling described below transform each original frame  $Y_k$  into an aligned frame  $D_k \in \mathbb{R}^{\beta n \times \beta m}$ :

It is our goal to compute a higher resolution version of a 3D depth map from aligned low resolution depth maps. When solving for the high resolution image we therefore have to resample the aligned high-resolution depth pixel grid of the target image. We performed experiments to determine the best resampling strategy. It turned out that a nearest neighbor sampling from the low resolution images is preferable over any type of interpolated sampling. Interpolation implicitly introduces unwanted blurring that leads to a less accurate reconstruction of high-frequency shape details in the superresolved result.

Our data term takes the following form:

$$E_{\text{data}}(X) = \sum_{k=1}^N \|W_k .* T_k .* (D_k - X)\|_2,$$

where  $*$  denotes element-wise multiplication.  $W_k \in \mathbb{R}^{\beta n \times \beta m}$  is a banded matrix that encodes the positions of  $D_k$  which one samples from during resampling on the high-resolution target grid.  $T_k \in \mathbb{R}^{\beta n \times \beta m}$  is a diagonal matrix containing 0 entries for all samples from  $D_k$  which are unreliable according to the ToF sensor’s readings, as described in the following:

Since a ToF camera relies on a sufficiently strong return of the emitted IR pulse to measure depth, certain scene characteristics lead to biased or totally wrong depth estimates. In consequence, if a surface reflects light away from the camera, or if it absorbs most of the light, depth measurements become unreliable. An example can be seen in Fig. 5, where the ball has problematic reflectance properties and the print on the box absorbs most of the light. Fortunately, a low amplitude of the returned light wavefront at each pixel (the SR 3000 camera we use gives access to an amplitude image) indicates the occurrence of such difficult situations and, thus amplitude serves as a form of confidence measure. We therefore use a thresholding approach, to detect and exclude low-confidence measurements with low amplitude. Technically this is implemented in the matrix  $T_k$  which multiplies unreliable samples by 0.

We would like to remark that the choice of error norm is critical to the quality of the final result. In essence, the norm decides at each high resolution depth pixel on how to choose a best target depth position given the depth values from all low resolution maps at that position. Previous depth superresolution methods, such as Schuon et al. [12] as well as many image superresolution methods, employ a  $\ell_1$ -norm. While a  $\ell_1$ -norm forces the depth value at a certain high-resolution grid point towards the median of registered low-resolution samples, an  $\ell_2$ -norm yields their mean. For very noisy data, the median is certainly reasonable since it rejects outliers. In contrast, the mean yields

a smoother surface reconstruction, since the averaging cancels out recording noise. From our experience using ToF data and our method, it is more beneficial to capitalize from the smoothing effect of a  $\ell_2$ -norm.

### 3.2. Regularization Term

The regularization or prior term guides the energy minimization to a plausible solution, and is therefore essential if data are sparse and noise-contaminated.

We seek a prior that brings out high frequency 3D shape features that were present in the original scenes in the up-sampled 3D geometry. At the same time the prior shall suppress noise in those regions that correspond to actually smooth 3D geometry. Finally we seek it to be convex.

All these properties can be enforced by designing a prior that favors certain distribution of the spatial gradient in the final depth map. On the one hand we want to preserve local maxima in the spatial gradient that correspond to high frequency features, e.g. depth edges. On the other hand, we want the overall distribution of the gradient to be smooth and relatively sparse which cancels out random noise.

One way to enforce this property is to resort to a sum-of-gradient-norms regularization term that can be computed efficiently, and that has also been used by previous image superresolution methods. However, the implementation of this regularizer for image superresolution often enforces sparseness on individual differences contributing to an overall finite difference approximation of the spatial gradient. For instance, the regularizer employed by Schuon et al. [12] essentially enforces sparseness on the elements of the approximated vector (i.e. sparseness on the individual finite differences). Although this prior manages to preserve high frequency detail to a certain extent, it completely fails in areas of smooth geometry where it creates a severe staircasing pattern (e.g. Fig. 2f). While small staircasing artifacts may not be visible if one works with intensity data, 3D reconstructions are severely affected.

We have therefore designed a new sum-of-norms prior that can be efficiently computed and that is tailored to produce high-quality 3D reconstructions. Let  $\nabla X_{u,v}$  be a combined vector of finite difference spatial gradient approximations at different scales at depth pixel position  $(u, v)$ . Then our regularization term reads:

$$E_{\text{regular}}(X) = \sum_{u,v} \|\nabla X_{u,v}\|_2 = \sum_{u,v} \left\| \begin{pmatrix} G_{u,v}(0, 1) \\ G_{u,v}(1, 0) \\ \vdots \\ G_{u,v}(l, m) \end{pmatrix} \right\|_2,$$

where each  $G_{u,v}(l, m)$  is a finite difference defined as follows

$$G_{u,v}(l, m) = \frac{X(u, v) - X(u + l, v + m)}{\sqrt{l^2 + m^2}}.$$

In our regularizer, we approximate the gradient with finite differences, but weight the various differences by the inverse Euclidean distances, yielding a rotation invariant approximation. Secondly we compute local gradient approximations at different scales and weight gradient approximations at lower levels of hierarchy (i.e. computed with a higher pixel position difference) lower. An important insight is that it is essential to compute the norm on all differences contributing to a local gradient approximation at different scales *simultaneously* and not on individual finite differences.

Since the  $(\ell_2)$ -norms of all combined gradient vectors in the above sum are positive, the sum has the effect of a  $\ell_1$ -regularization [1] on the entire set of gradient magnitudes: enforcing sparseness, i.e. drive most gradients to zero and hence smooth the result in noisy regions, but allow high-frequency detail to prevail. By combining distance-weighted gradient approximations at different scales we thus implicitly achieve feature preserving smoothing in a computationally efficient and convex way.

Given the data and regularization terms defined in the previous sections, we can now formulate the complete LidarBoost energy function as

$$\sum_{k=1}^K \|T_k \cdot W_k \cdot (D_k - X)\|_2 + \lambda \sum_{u,v} \|\nabla X_{u,v}\|_2,$$

where  $\lambda$  is the trade-off parameter between enforcement of data similarity and smoothness. As one can see in Fig. 2g, our approach produces high quality superresolved geometry which exhibits clear 3D features and only marginal noise in smooth areas.

## 4. Results

To explore the capabilities of the new approach, we tested it on synthetic and real sequences captured with a Swissranger SR3000 camera ( $176 \times 144$  depth pixel resolution). We also compared LidarBoost to two alternative approaches from the literature. First we compare against an image-based superresolution method applied to depth data (IBSR), in particular we use the approach by Farsiu et al. [3] as proposed by Schuon et al. [12]. We apply the publicly available implementation of Farsiu's approach and choose the following parameters:  $\lambda = 0.04$ ,  $N = 50$ ,  $\alpha = 0.7$ ,  $\beta = 1$ ,  $P = 5$ , and a Gaussian  $3 \times 3$  PSF with standard variance (see original paper for details). The computation time was below two minutes for all scenes.

Second, on the real scenes only, we compare against color and depth fusion method, namely the method by Diebel and Thrun [2]. We ran all method with several parametrizations and show only the best results for each method in the respective scenes.

LidarBoost was implemented using the cvx modeling framework for disciplined convex optimization [4]. This

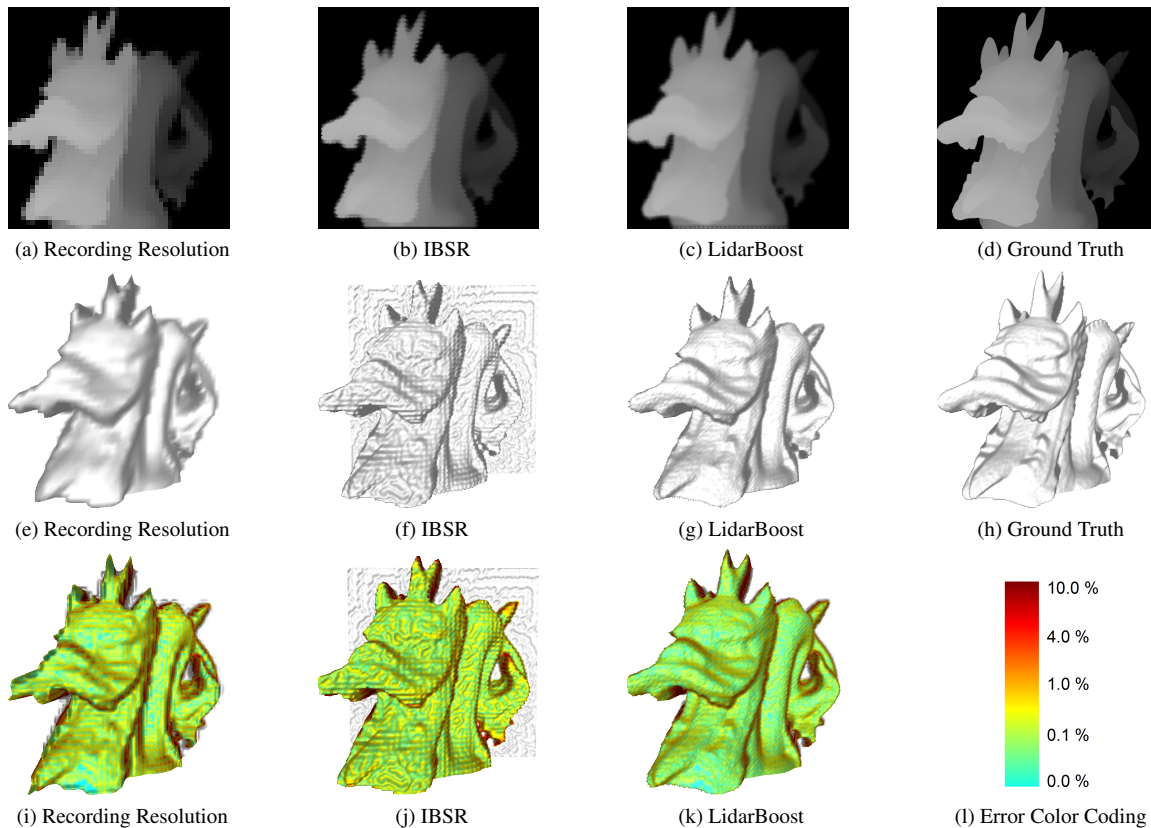


Figure 2: Synthetic test set without noise ( $4\times$  upsampling): The first row depicts the depth maps, from which a 3D geometry has been rendered as shown in the second row. The third row shows a rendering, with color coded rMSE. IBSR recovers the overall structure, but exhibits a noise pattern. LidarBoost recovers the structure almost perfectly and yields a smooth surface.

framework transforms the problem into Second-Order-Cone-Program (SOCP) and solves it using a generic solver. Due to the size of the transformed problem, which easily exceeded a million variables, we subsequently compute solutions for images patches of  $20 \times 20$  low-resolution pixels and stitch the results using two-pixel overlap (similar to primal decomposition with one iteration). Computation time for the synthetic scenes (9 patches) was about five minutes and for the real scenes (28 - 48 patches) up to two hours.

**Synthetic Scene - No Noise Added** A first comparison is performed on synthetic images of the Stanford Graphics Lab’s dragon model created with 3D Studio Max. Synthetic ground truth depth maps of resolution  $400 \times 400$  were rendered and downsampled by factor 8 (using a uniform  $8 \times 8$  kernel) to simulate low resolution input depth maps. In total,  $N = 10$  low resolution input images from slightly displaced viewpoints were created. One such input depth maps is shown in Fig. 2a, compared to the ground truth shown in Fig. 2d. Figs. 2b and 2c show the four times superresolved results computed by applying IBSR and LidarBoost. Below each depth map, we show renderings of the corresponding

3D geometry (obtained by reprojection into 3D) since depth maps only do not properly visualize the true gain in quality and tend to mask unwanted artifacts.

The method by Schuon et al. successfully brings out the outline of certain shape detail that was not visible in individual input frames, Fig. 2f, such as individual toes and sharp boundaries. However, the results are clearly contaminated by the previously discussed staircase pattern (Sect. 3.2). In comparison, LidarBoost (Fig. 2g, with  $\lambda = 0.04$ ) extracts more detail (e.g. the eye holes and the third small horn of the dragon) and at the same time successfully eradicates measurement noise without introducing a disturbing pattern.

On synthetic data we can also perform quantitative comparisons against ground truth and compute the relative mean square error. It is relative, because the MSE result was divided by the number of pixels considered to keep numbers reasonable. A two times downsampled version of a reference  $400 \times 400$  depth depth map forms the ground truth - to make resolutions match. One low resolution depth map has been upsampled four times using a nearest neighbor

	No Noise <i>var</i> = 0	Medium Noise <i>var</i> = 0.7	Stark Noise <i>var</i> = 5
LR	157.6	161.7	203.9
IBSR	83.8	89.9	127.0
LidarBoost	70.6	72.5	82.9

Table 1: Relative MSE comparison on synthetic data: LidarBoost throughout outperforms all other methods and shows less sensitivity towards noise than IBSR

approach to establish a baseline comparison. LidarBoost clearly outperforms the IBSR method and leads to significant improvements over a single low-resolution depth map. Figs. 2i - 2k show a color-coded rendering of the error distribution (rMSE in percent of longest bounding box dimension of synthetic object) over the model surface using the color scheme shown in Fig. 2l (green=low error, red=large error).

Both methods struggle on edges, which comes to no surprise, as the sub-pixel exact location for a steep edge is hard to guess. Despite a potentially small mis-localisation, LidarBoost still recovers depth edges more reliably than the comparison method. Also, the pattern introduced by IBSR leads to much stronger errors in the interior regions than with LidarBoost.

**Synthetic Scene - Stark Noise Added** Depth images are inherently noisy, therefore the algorithms need to be evaluated on such data. To simulate the effect of measurement noise introduced by real ToF cameras, we repeated the experiment from the previous section, but added Gaussian noise along the measurement ray directions following the sensor characterization proposed by Kim et al. [6]. In the simulated data, depth values range from 0 to 182. Although in scenes with a larger depth range a depth-dependency in noise can be expected, for our test scene with limited range we use a constant variance. We tested both noise with variance of 0.7 and 5.0. In both cases the LidarBoost results are clearly superior to IBSR, and for space reasons we decided to discuss here only the case of strong noise contamination (the low noise case is shown in the additional material).

For the stark noise case, in a single low resolution input frame (Fig. 3a) all fine surface detail vanished and it is even hard to recognize the object’s shape as a whole. While IBSR recovers a decent level of shape detail (Fig. 3b), severe staircasing becomes visible on the geometry and the result is distorted by the random pattern discussed before. In contrast, in particular under these extreme conditions, LidarBoost recovers clearly more detail (even traces of the dragon’s pattern on the back, as well as the dragon’s teeth) and maintains truly smooth geometry in actually smooth areas. The color-coded error rendering confirms that under these challenging conditions the advantage of using LidarBoost relative to IBSR is even stronger, Figs. 3i - 3k.

**Synthetic Scenes - Quantitative Comparison** Looking at all synthetic test data sets, the overall trend in rMSE error confirms the visual observations (Tab. 1). In all noise cases our algorithm performs clearly better than the reference approach and clearly improves over the quality of a single low resolution frame. Overall, with increasing noise the performance of IBSR worsens more drastically than our method’s results.

**Parameter Selection** Both LidarBoost and IBSR use a regularization term with a tunable trade-off parameter  $\lambda$ . Fig. 4 plots  $\lambda$  against the rMSE obtained with both IBSR and LidarBoost, as evaluated on the dragon data set with stark noise. The reconstruction quality of the former shows a strong dependency on  $\lambda$ , and the rMSE is in general much higher than that for LidarBoost. In contrast, the rMSE of LidarBoost is consistently lower and rather stable. Therefore  $\lambda$  requires less tweaking which renders LidarBoost highly applicable. The same observation was made for data sets with no noise and stark noise (additional evidence is given in the additional materials).

**Real Scene - Collection of Objects** Two real scenes were recorded using a Swissranger SR 3000 depth camera. We recorded  $N = 15$  frames each with 30 ms integration time. The camera was displaced in-between shots using rotation only, where the maximum displacement from end to end was below 30 pixels for the first and below 15 pixels for the second scene. The SR 3000 records at  $176 \times 144$  pixel resolution, but we cropped the frames in either case to the region of interest, which for the collection of objects scene (Fig. 5a) resulted in a  $106 \times 64$  frame size, and for the second scene (Fig. 1a) in a  $126 \times 89$  frame size.

For this scene, the low resolution input (one being shown in Fig. 5b) conveys the overall geometry, but fine details such as the small holes of the laundry basket and the cup’s handle are hard to tell. Also, the occlusion edges are very rough and aliased. Furthermore smooth surfaces, such as the ball’s or basket’s surface are perturbed by noise.

IBSR’s reconstruction enhances the fine details, but also introduces the previously discussed staircase pattern. In contrast, LidarBoost (running with  $\lambda = 7$ ) also does feature these details, while yielding a noise free, smooth surface. This result also shows the effectiveness of our am-

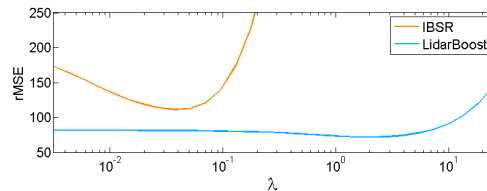


Figure 4: Optimal choice of regularization trade-off parameter  $\lambda$ : For the noisy test sets the resulting rMSE has been plotted against varying  $\lambda$ . IBSR is sensitive towards  $\lambda$  with a constant optimum at 0.04. In contrast LidarBoost is robust on a wide range of choices.

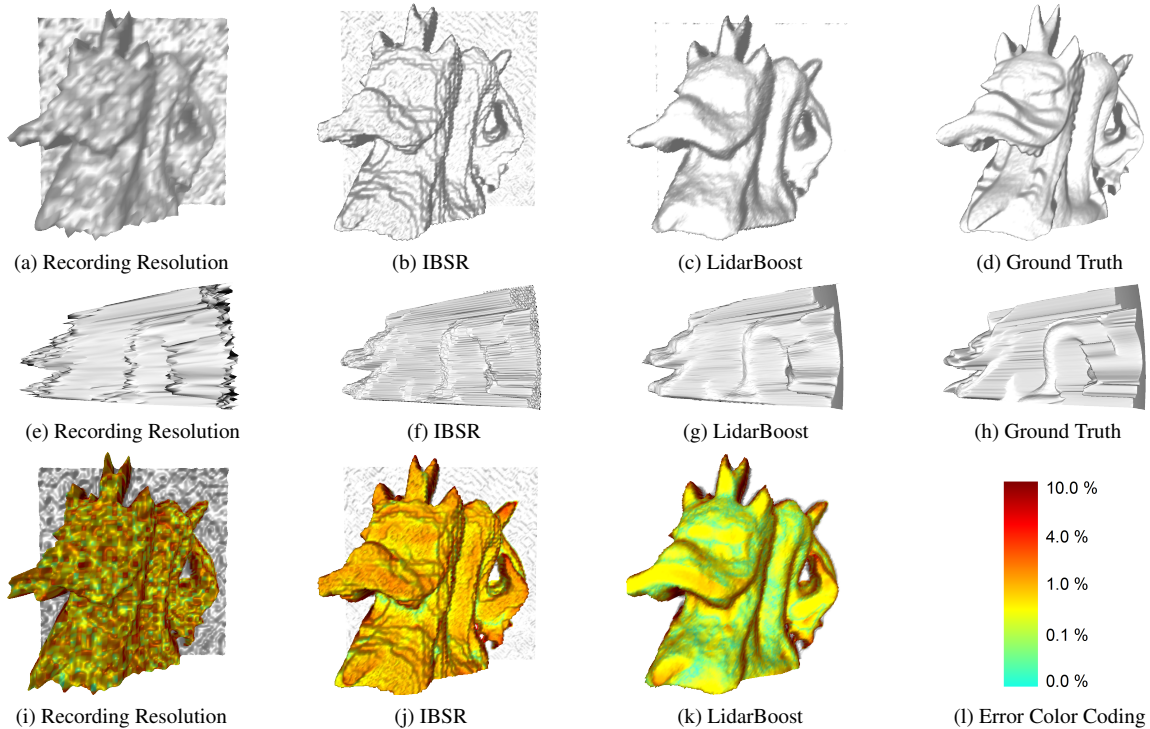


Figure 3: Synthetic test set with stark noise (Variance of 5.0,  $4\times$  upsampling) - First row: rendered 3D geometry in frontal view, LidarBoost shows shows best upsampling result. Middle row: Also in a lateral view it is apparent that LidarBoost’s reconstruction is closest to ground truth. Bottom row: LidarBoost clearly produces the lowest reconstruction error.

plitude thresholding approach. Parts of the cardboard are painted in black, leading to low reflectivity. Fig. 5c shows the amplitude image with measurements below the experimentally determined thresholds being color coded in red. By assigning such pixels a weight of 0 via  $T_k$ , LidarBoost reconstructs the true surface (5f) of the box. Please also note that two stripes of reflective material on the soccer ball caused slight reconstruction errors since almost no light was reflected to the camera. In this particular case our confidence weighting could not fill the holes since the tiny area of correctly capture depth on the rim pulls the final surface slightly inward.

Since we also took a photograph of the real scene, we can also compare to the method by Diebel et al. (Fig. 5g) which yields a smooth reconstruction, but struggles with fine details such as the basket’s bars, and oversmooths depth edges that don’t coincide with intensity edges. Furthermore the method erroneously transforms intensity texture into geometric patterns, in particular in the checkerboard structure on the background and in the pattern on the ball’s surface.

**Real Scene - Wedges and Panels** The second real scene recorded with the Swissranger was purposefully designed to contain wedges with thin fine edges, and many sharp occlusion boundaries (Fig. 1a). The same camera settings as in the previous test were used and  $N = 15$  low resolution

frames were captured. This scene nicely demonstrates the effectiveness of superresolution. While in the low resolution image (Fig. 1b), occlusion edges clearly show a stair-casing aliasing pattern, both IBSR and LidarBoost recover sharper edges. However, in Schuon’s result there is still a little bit of jaggedness around the occlusion edges and, as in previous results, there is a strong aliasing pattern in regions of smooth geometry (Fig. 1c). In contrast, LidarBoost (with  $\lambda = 6$ ) creates crisp edges with no aliasing, and faithfully recovers smooth areas (Fig. 1d). In addition, LidarBoost does a much better job in recovering different depth layers that are visible through the small holes in the left panel (marked in red in Fig. 1d).

Diebel et al.’s method does well in recovering the layers, but in contrast to our method exhibits problems on several edges. Many edges on the wedges appear rounded or are still aliased (particularly on the right most wedge).

## 5. Conclusion

We presented a new 3D depth superresolution approach that enables us to take detailed high resolution surface recordings of 3D objects even with low resolution and heavily noisy time-of-flight cameras. The core of our method is an energy minimization framework that explicitly takes into account the characteristic of the sensor and the specific re-

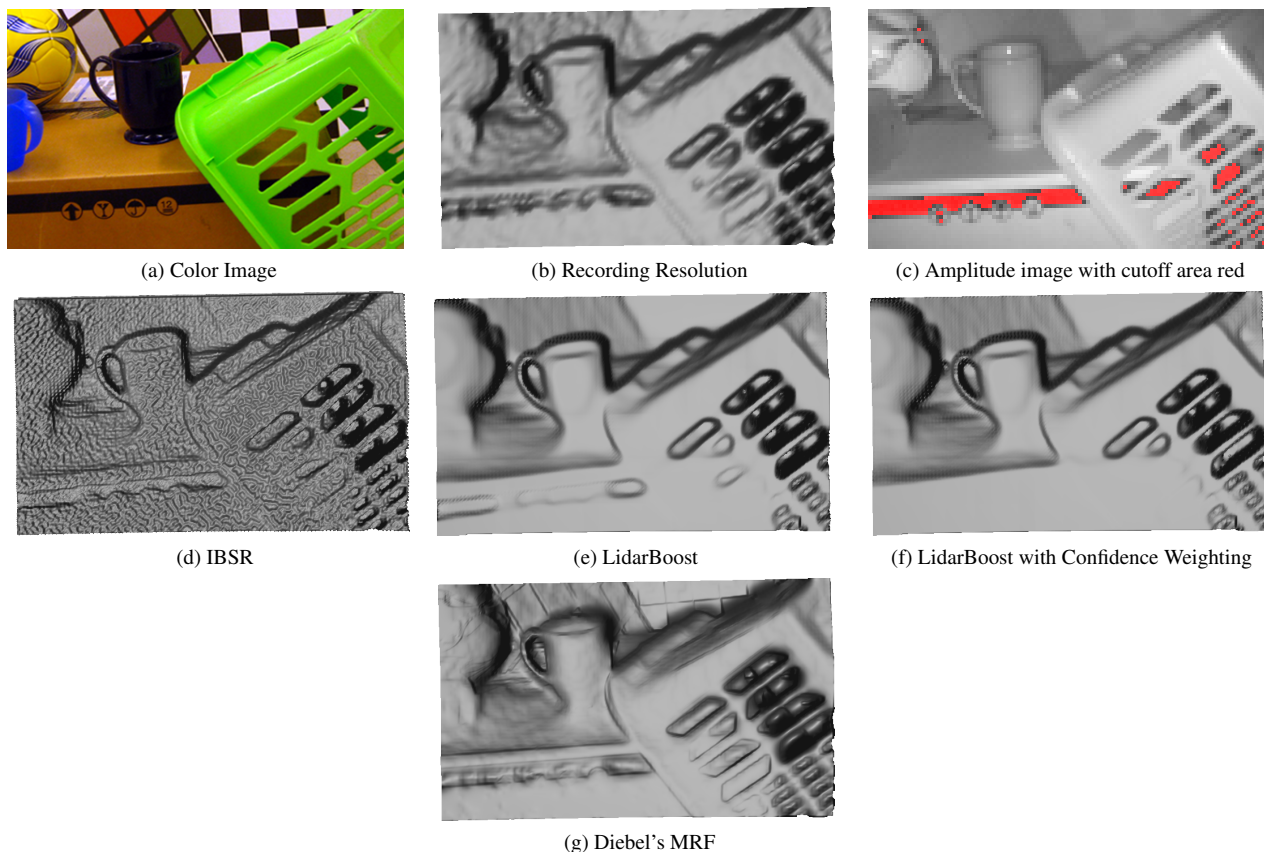


Figure 5: Real scene - collection of objects (a): One of several low-resolution depth maps with an SR3000 ToF cam is shown in (b). IBSR (d) produces an erroneous pattern, whereas LidarBoost (e) correctly recovers high-frequency detail and smooth geometry. When the reflectivity of the materials is really low, the low resolution recordings may contain errors (such as in the red areas in (c)). LidarBoost with activated confidence weighting (f) can correct for such reconstruction errors. Diebel's MRF method (g) yields oversmoothing on many depth edges and transforms intensity patterns into geometry patterns (e.g. checkerboard).

quirements to superresolution in 3D space. We have shown both quantitatively and qualitatively that our algorithm produces high-quality results and outperforms related methods from the literature.

## References

- [1] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.
- [2] J. Diebel and S. Thrun. An application of markov random fields to range sensing. In *Advances in Neural Information Processing Systems 18*, pages 291–298. 2006.
- [3] S. Farsiu, M. D. Robinson, M. Elad, and P. Milanfar. Fast and robust multiframe super resolution. *IEEE Transactions on Image Processing*, 13(10):1327–1344, Oct. 2004.
- [4] M. C. Grant, S. P. Boyd, and Y. Ye. CVX: Matlab Software for Disciplined Convex Programming. <http://stanford.edu/boyd/cvx>, 2008.
- [5] Y. Kil, B. Mederos, and N. Amenta. Laser scanner super-resolution. *Eurographics Symposium on Point-Based Graphics*, 2006.
- [6] Y. Kim, D. Chan, C. Theobalt, and S. Thrun. Design and Calibration of a Multi-view TOF Sensor Fusion System. *IEEE CVPR Workshop on Time-of-flight Computer Vision 2008*, 2008.
- [7] J. Kopf, M. Cohen, D. Lischinski, and M. Uyttendaele. Joint bilateral upsampling. *ACM TOG*, 26(3), 2007.
- [8] B. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *International Joint Conference on Artificial Intelligence*, volume 3, 1981.
- [9] S. Park, M. Park, and M. Kang. Super-resolution image reconstruction: a technical overview. *Signal Processing Magazine, IEEE*, 20(3):21–36, 2003.
- [10] A. Rajagopalan, A. Bhavsar, F. Wallhoff, and G. Rigoll. Resolution Enhancement of PMD Range Maps. *Lecture Notes in Computer Science*, 5096:304–313, 2008.
- [11] G. Rosenbush, T. Hong, and R. Eastman. Super-resolution enhancement of flash LADAR range data. *Proceedings of SPIE*, 6736:673614, 2007.
- [12] S. Schuon, C. Theobalt, J. Davis, and S. Thrun. High-quality scanning using time-of-flight depth superresolution. *CVPR TOF Workshop 2008*, 2008.
- [13] S. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski. A comparison and evaluation of multi-view stereo reconstruction algorithms. In *IEEE CVPR*, pages 519–528, 2006.
- [14] Q. Yang, R. Yang, J. Davis, and D. Nistér. Spatial-depth super resolution for range images. In *IEEE CVPR*, 2007.