Open access • Proceedings Article • DOI:10.1109/CVPR.2010.5540082

# 3D shape scanning with a time-of-flight camera — Source link

Yan Cui, Sebastian Schuon, Derek Y. Chan, Sebastian Thrun ...+1 more authors

**Institutions:** Stanford University

**Published on:** 13 Jun 2010 - Computer Vision and Pattern Recognition

**Topics:** Time-of-flight camera, Image sensor, Noise (video) and Image resolution

# 3D Shape Scanning with a Time-of-Flight Camera

Yan Cui[1,3], Sebastian Schuon[2], Derek Chan[2], Sebastian Thrun[2], Christian Theobalt[1]

[1]MPI Informatik [2]Stanford University [3]Augmented Vision, DFKI

Yan.Cui@dfki.de {schuon,ddc,thrun}@cs.stanford.edu theobalt@mpi-inf.mpg.de

## Abstract

*We describe a method for 3D object scanning by aligning depth scans that were taken from around an object with a time-of-flight camera. These ToF cameras can measure depth scans at video rate. Due to comparably simple technology they bear potential for low cost production in big volumes. Our easy-to-use, cost-effective scanning solution based on such a sensor could make 3D scanning technology more accessible to everyday users. The algorithmic challenge we face is that the sensor's level of random noise is substantial and there is a non-trivial systematic bias. In this paper we show the surprising result that 3D scans of reasonable quality can also be obtained with a sensor of such low data quality. Established filtering and scan alignment techniques from the literature fail to achieve this goal. In contrast, our algorithm is based on a new combination of a 3D superresolution method with a probabilistic scan alignment approach that explicitly takes into account the sensor's noise characteristics.*

## 1. Introduction

Nowadays, 3D geometry models of real world objects are essential in many application scenarios, such as design and virtual prototyping, quality assurance or applications in visual media, such as games, virtual worlds and movie special effects - to name just a few. Existing 3D shape scanning technology is often based on rather specialized and complex sensors, such as structured light camera/projector systems or laser range finders, Sect. 2. Even though they produce data of high quality, they are quite expensive and often require expert knowledge for their operation. It is thus no wonder that semi-professional or everyday users have usually no access to such technology. On the other hand, if easy-to-operate and cheap 3D scanners were more amenable, 3D shape models could turn into a much more widely used asset, just as image and video data are today. This could open the door for many new applications, for instance in community web platforms or online shopping.

In this paper, we therefore propose a new easy-to-use 3D object scanning approach based a time-of-flight (ToF) 3D camera. A ToF camera has a variety of advantages over alternative 3D scanning technology: It can measure 3D depth maps at video rate and thus lends itself for integration into a fast object scanner. It is an active sensor that measures the travel time of infrared light, and therefore it does not interfere with the scene in the visual spectrum [17, 16]. Its core components are a CMOS chip and an infrared light source which bears the potential for low cost production in big volumes. Finally, its practical operation is no different from a video camera and can thus be easily performed by everyday users.

The biggest algorithmic challenge we face when putting this idea into practice is also the reason why ToF cameras have not yet taken over the 3D scanning market: ToF sensors have a very low X/Y resolution, an adverse random noise behavior, and a notable systematic measurement bias [1]. After a first look at the data quality of a single ToF depth scan, *e.g*. Fig. 1b, one may be tempted to not even try to use such a camera for shape scanning. However, in this paper we show that an appropriate combination of ToF specific resolution enhancement and scan alignment enables us to combine ToF scans taken from around an object into 3D shape models of reasonable quality, Fig. 1c. Shape acquisition is rather flexible and can be performed by rotation of an object in front of the camera or by hand-held motion of the camera around the object. Our main contributions therefore are, Sect. 3: 1) A 3D shape scanning approach based on a Time-of-Flight Camera. 2) A ToF-specific probabilistic procedure for simultaneous non-rigid alignment of multiple depth scans. 3) The integration of a ToF 3D superresolution approach with this alignment procedure into a complete ToF shape scanning approach. We will show that the combination of these steps and the explicit consideration of the camera's noise behavior is essential to make this possible. We tested our algorithm on a variety of objects and show that it compares favorably to laser scanning in visual and quantitative comparison, Sect. 4.


(a) Color Image (b) Raw Scan (c) Proposed Method (d) Laser Scan (e) Error Plot (f) Color Coding Legend

Figure 1: Antique head (a); our algorithm computes a 3D model of reasonable quality (c) despite severe errors in the raw ToF data (b). Reconstruction error (e) compared to a laser scan (d) shows that no circumstance the error was larger than 2.5 cm, while for most of the surface it was below 1.0 cm.(Note: raw aligned scans, no hole filling done)

## 2. Related Work

Most commercial systems for 3D shape scanning are based on structured light projection or laser stripe projection, please refer to [18, 10] for a recent overview. Passive image-based methods have also been successfully used for 3D shape reconstruction [24]. To build a complete 3D model, it is commnon to align several scans taken from different viewpoints (usually under very controlled motion). In contrast to ToF cameras, the above mentioned sensors provide rather clean data of relatively low random noise and systematic error. On such data, local rigid alignment techniques, such as Iterative Closest Points (ICP) and its variants [5] or global rigid alignment techniques, *e.g.* [4, 3, 11] can be used to register the scans against each other. Finally, a scan merging procedure, such as [8] can be applied to build a single 3D mesh. Hand-held scanners based on the above technologies have been proposed where the camera can be freely moved around an object (or vice versa), *e.g.* [22]. Our work supports both hand-held scanning and scanning under controlled motion, *e.g.* with a turntable. A related idea to build a relatively simple 3D scanner has been proposed by Bouguet et al. [6] who measure 3D shape by recording a shadow cast by a rod moved over the object. However, freely moving the scanner around the object is not easy with this approach.

So far, time-of-flight (ToF) cameras [17, 16] have rarely been explored as sensors for 3D object scanning [12, 25], even though they have a variety of advantages over the above technologies (see Sect. 3 for details). This is mainly due to the challenging noise and bias characteristics [1] which renders direct application of established filtering and alignment approaches infeasible. Some previous work proposes pre-calibration approaches to compensate for instance bias effects in the cameras [19]. Others tried to attack ToF camera deficiencies by combining them with normal color cameras, *e.g.* [26, 2]. In contrast, in this paper we show that reliable shape capture is feasible with ToF cameras alone. To this end we capitalize on recent time-of-flight superresolution methods [21, 23]. Related to these methods is the method by Kil et al. [14] for 3D superresolution with a laser scanner. The former approaches are designed for the more challenging noise characteristics of ToF cameras. In addition, in our algorithm the systematic camera error is compensated by a new global non-rigid scan alignment approach. Some previous work already deals with global non-rigid shape alignment [7], but not under consideration of ToF specifics. Related to our work is also research on surface reconstruction from noisy, but already aligned, scans [13, 9]. Our approach differs in that it extends previous work on probabilistic non-rigid alignment of pairs of scans [20] into a global method. Suitable rigid and non-rigid scan alignment is achieved by explicitly incorporating ToF specific noise characteristics.

## 3. Our Algorithm

Our goal is to build the, to our knowledge, first 3D shape scanner based on a Time-of-Flight camera that can be used in hand-held and turntable scanning mode. We use a MESA Swissranger SR4000 as ToF sensor. In a nutshell, it emits infrared light into the scene and at each pixel measures the return time of the reflected light from which it determines the depth of the pixel. More about the phase-shift based internal measurement principle of the SR4000 can be found, for instance, in [17, 16].

The time-of-flight sensor has a variety of conceptual advantages over previously used sensor techniques for shape scanning: it captures full frame depth at video rate, *i.e.* it does not need to subsequently scan scene points for a single depth map (like a laser scanner), it does not rely on time multiplexing like structured light scanners ( even though internally the ToF camera performs several measurements; for normal motion at normal speed this effect is negligi-

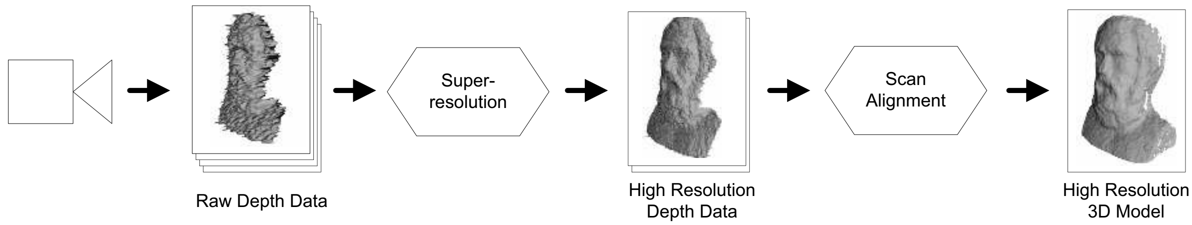



Figure 2: Outline of our processing pipeline.

ble). Additionally, its measurement quality is largely independent from scene texture and it does not interfere with the scene in the visual spectrum enabling simultaneous texture recording. Finally, it is based on comparably simple CMOS technology enabling low cost manufacturing in large numbers.

During acquisition the camera is either moved by hand in an arc around the object, or the object is turned in front of the static camera, *e.g.* using a turntable. The object should stay roughly in the center of the field of view and the distance of the camera to the object is kept approximately constant, Fig 3. This way, a sequence of $i = 1, \ldots, N_c$ depth maps $D_i$ is obtained from $N_c$ subsequent positions along the camera path, each of which can be described by a tuple of a rotation matrix and a translation vector with respect to a global coordinate system $(R_i, t_i)$.

Unfortunately, the previously mentioned advantages of the ToF sensor come at the price of very low X/Y sensor resolution ($N_x = 176 \times N_y = 144$ pixels for the SR4000), random depth noise with significant standard deviation and a substantial systematic measurement bias that distorts the depth maps [1, 15], see Figs. 1, 2 and 8 for examples of raw ToF depth scans. Our new approach presented in this paper enables us to combine and align these rather low quality depth scans. The result is a 3D model of substantially higher quality than a single depth scan would suggest. It's algorithmic core and the main novelty is a combination of 3D ToF superresolution with a new ToF-specific probabilistic method for simultaneous rigid and non-rigid alignment of multiple depth scans. Our algorithm comprises the following steps, Fig. 2:

1) Initial estimates for the scan poses $(R_i, t_i)$ are found by using the Voodoo camera tracker [1] on the amplitude images provided by the camera for each depth map $D_i$.

2) We subdivide the set of $N_c$ depth images into a set of $\ell = 1, \ldots, K$ chunks of depth images, Fig. 3. Each chunk $\mathcal{C}_\ell = (D_{\rho(\ell)}, \ldots, D_{\rho(\ell)+C})$ comprises $C$ subsequent depth images starting from a frame index $\rho(\ell)$. To each chunk of depth images a ToF superresolution approach is applied, yielding $K$ new depth maps $H_\ell$ with much higher X/Y resolution, Section 3.1.

3) the $H_\ell$ are converted to 3D geometry $Y_\ell$ and aligned by means of a probabilistic alignment approach that not only recovers the rigid scan alignment parameters, but also compensates for non-rigid bias-induced deformations, Sects. 3.2 and 3.3. Please note that we don't perform explicit scan merging, *e.g.* to build a single surface mesh. Any surface reconstruction or scan merging strategy from the literature can be applied after our alignment.

[1] Voodo camera tracker: www.digilab.unihannover. de/index.html

## 3.1. Superresolution

To each chunk of frames $\mathcal{C}_\ell$, we apply the LidarBoost ToF superresolution approach [23] which yields a high-resolution depth map aligned to the center frame of the chunk. In the following we briefly describe the core concepts of LidarBoost and refer the reader to [23] for more detail. First, all depth maps in the chunk are aligned to the center frame using optical flow. This is sufficiently accurate since the maximum viewpoint displacements throughout the entire chunk are typically one to two depth pixels. LidarBoost extracts a high-resolution denoised center depth map $H_\ell$, Fig. 2, from the aligned low resolution depth map by solving an optimization problem of the form:

$$\min_{H_\ell} E_{\text{data}}(L_{\rho(\ell)}, \ldots, L_{\rho(\ell)+C}, H_\ell) + E_{\text{reg}}(H_\ell) \ . \qquad (1)$$

Here, $L_{\rho(\ell)}, \ldots, L_{\rho(\ell)+C}$ are the raw depth maps aligned to the center one. $E_{\text{data}}$ measures the agreement of $H_\ell$ with the aligned low resolution maps; unreliable depth pixels with low amplitude are discarded. $E_{\text{reg}}$ is a feature-preserving smoothing regularizer tailored to ToF data. We upsample the depth data by factor 4 in both X and Y resolution. The

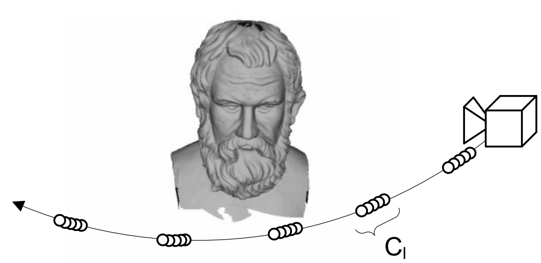



Figure 3: A typical camera path: The dotted segments are the frame chunks $C_\ell$ from which superresolved depth scans are computed.

$H_\ell$ are straightforwardly converted into 3D point clouds $Y_\ell$ by reprojection into space using the ToF camera's intrinsic parameters (calibrated off-line).

### 3.2. Systematic Bias

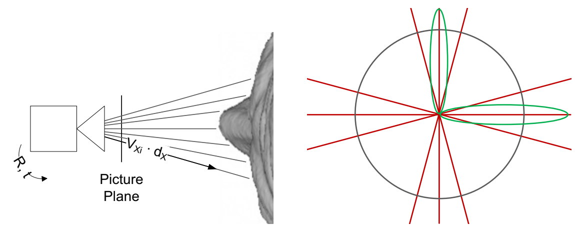
While the random noise is effectively reduced by the superresolution approach, the ToF data's systematic bias leads to non-rigid ToF scan distortions and therefore needs special attention. Let $x_i$, $i = 1, \ldots, N_x \times N_y$ be a 3D point measured by the depth camera (*i.e.* the point in space after reprojection of the depth pixel), and $V_i$ be the the direction of the camera ray towards the point. Then the bias can be modeled as a systematic offset $d_i$ along this ray which makes the camera measure the value $x_i = \tilde{x_i} + V_i d_i$ rather than the true 3D point $\tilde{x_i}$ (Fig. 4a). Previous studies have shown that the depth bias is pixel dependent and dependent on many factors, including the camera's integration time, scene reflectance, surface orientation, and distance [1, 15].

Accounting for all such dependencies in our framework would render the problem intractable. We therefore make a few simplifying assumptions. First, in practice the bias dependency on reflectance, surface orientation, and integration (since it stays constant for a scan) can be neglected. Second, the depth range covered by the scanned object is usually limited and the distance of the camera to the object remains fairly constant. Therefore, we ignore the depth dependency of the bias. Finally, when averaging hundreds of depth frames of a flat wall, the resulting 3D model typically shows a radially symmetric deviation from the plane, with increasing curvature (bias) the further one is away from the depth image center. We therefore assume that all depth pixels with the same radial distance from the image center have the same bias, and the bias increases with the radius. Since the $H_\ell$ are computed from closely spaced low resolution ToF scans, we assume the above bias characteristics also applies to them. The set of radially symmetric bias values therefore is parameterized as $(d_1, \ldots, d_O)$ with $O$ being half the maximum number of pixels on the diagonal of all $H_\ell$.

### 3.3. Probabilistic Simultaneous Scan Alignment

The last and most important step of ToF scan reconstruction is a probabilistic global alignment approach that solves for the rigid alignment $(R_\ell, t_\ell)$ of all high-resolution 3D point clouds $Y_\ell$, as well as the systematic bias values $(d_1, \ldots, d_O)$. Rather than pre-compensating the bias, as in [15], we explicitly model the set of bias variables as unknowns of the alignment procedure. This enables us to accommodate for the potential scene dependency of the bias to a certain degree while keeping the number of variables in reasonable bounds.

Our algorithm is inspired by the non-rigid registration approach for pairs of scans described in [20]. We extend




(a) (b)

Figure 4: (a) During multi-scan alignment, the motion if the points in the scan is parameterized by a rigid component $(R, t)$ and a non-rigid warp along the viewing ray direction (representing measurement bias). (b) Point sampling strategy during alignment (in pixel domain). Red line is the sparse sampling strategy for the reference set $X$; the green region is the sampling region for the $Y_\ell$.

their ideas to our setting and develop an approach for simultaneous rigid and non-rigid multi-scan alignment that incorporates knowledge about the ToF bias characteristics. ToF scan registration is formulated as a maximum-likelihood estimation problem. We choose any one of the high-resolution 3D point clouds as the reference 3D point set, henceforth termed $X = \{x_n \mid n = 1, \ldots, N_r\}$. Each of the remaining high-resolution 3D point clouds, $Y_\ell$, is simultaneously aligned to the reference point cloud. For ease of explanation, in the following we describe the alignment process for a single point cloud, henceforth $Y = \{y_m \mid m = 1, \ldots, N_d\}$, before we show how this process is applied to all scans.

On each point in $Y$, a multi-variate Gaussian is centered. All Gaussians share the same isotropic covariance matrix $\sigma^2 I$, $I$ being a $3 \times 3$ identity matrix and $\sigma^2$ the variance in all directions. Hence the whole point set can be considered a Gaussian Mixture Model (GMM) with density:

$$p(x) = \sum_{m=1}^{N_d} \frac{1}{N_d} p(x|m) \quad with \quad x|m \propto N(y_m, \sigma^2 I) \; . \tag{2}$$

Alignment of $Y$ to $X$ is performed by maximizing the likelihood function. In our setting, the motion of $Y$ towards $X$ is parameterized by the rigid motion component $(R, t)$, as well as the bias motion component, *i.e.* a translation of each point $y_m$ along the local viewing ray direction $V_m$ by the bias factor $d_m$, Figure 4a. Please note that while the rigid component for $X$ remains fixed, the bias-induced non-rigid deformation is also applied to $X$ when searching for optimal alignment. The maximum likelihood solution for $(R, t)$ and $d_1, \ldots, d_O$ is found by minimizing the negative log-likelihood which yields the following energy func-

- Initialize: $R_\ell, t_\ell\ \forall \ell = 1,\dots,K$ from structure from motion (Sect. 3); $(d_1,\dots,d_O) = 0$
- Repeat until no further improvement or max. iterations :
  - Update variance $\sigma^2$ according to Eq. (4)
  - Rigid registration, for all pairs $Y_\ell$,$X$ (case I):
    1. E-step: Compute $P$ for pair $Y_\ell$,$X$
    2. M-step: Solve for $R_\ell$, $t_\ell$ by minimizing Eq. (7) as linear least squares system
  - Non-rigid registration, for all pairs $Y_\ell$,$X$ (case II):
    1. E-step: Compute $P$ for all pairs $Y_\ell$, $X$
    2. M-step: Solve for $(d_1,\dots,d_O)$ by combining $O \cdot K$ linear equations of the form $\frac{\partial(Q_{non-rigid}(d_1,\dots,d_O))}{\partial d_i} = 0$ into a joint linear least squares system (*i.e.* combine the equations for all $\ell = 1,\dots,K$)

Figure 5: Probabilistic simultaneous scan alignment

tional:

$$E(R,t,d_1,\dots d_O) = \\ -\sum_{n=1}^{N_r} \log \sum_{m=1}^{N_d} \exp\left(-\frac{1}{2}\left\|\frac{x_n+V_{x_n}d_n-(R(y_m+V_{y_m}d_m)+t)}{\sigma}\right\|^2\right) \\ +\lambda\left\|(d_1-d_2,\dots,d_{O-1}-d_O)\right\|^2 . \quad (3)$$

The variance $\sigma^2$ of the mixture components is estimated using

$$\sigma^2 = \frac{1}{N_r N_d}\sum_{n=1}^{N_r}\sum_{m=1}^{N_d}\left\|x_n-(Ry_m+t)\right\|^2 . \quad (4)$$

Note that our energy functional contains a regularization term weighted by $\lambda$ ($\lambda = 3$ in all our experiments) which ensures a smooth distributions of the bias values. In accordance with Sect. 3.2 the term favors monotonous bias distributions with increasing radius from the image center.

We use an iterative Expectation Maximization (EM) like procedure to find a maximum likelihood solution of Eq. (3). During the E-step the best alignment parameters from the previous iteration are used to compute an estimate of the posterior $p^{old}(m|x_n)$ of mixture components by using Bayes theorem. During the M-step, new alignment parameter values are found by minimizing the negative log-likelihood function, or more specifically, its upper bound $Q$ which evaluates to:

$$Q(R,t,d_1,\dots,d_O) = \\ \sum_{n=1}^{N_r}\sum_{m=1}^{N_d} P^{old}(m|x_n)\frac{\|x_n+V_{x_n}d_n-(R(y_m+V_{y_m}d_m)+t)\|^2}{2\sigma^2} \\ +\lambda\left\|(d_1-d_2,\dots,d_{O-1}-d_O)\right\|^2 . \quad (5)$$

The above EM procedure converges to a local minimum of the negative log-likelihood function. Please note that the variances $\sigma^2$ are continuously recomputed which is similar to an annealing procedure in which the support of the Gaussians is reduced when point sets get closer.

Experimentally, we could verify that a simultaneous optimization of all alignment parameters often fails to converge to a suitable minimum. Instead, we propose to alternate between optimizing for $(R,t)$ with fixed $(d_1,\dots,d_O)$ (case I), and optimizing for $(d_1,\dots,d_O)$ with fixed $(R,t)$ (case II).

In case I, $p^{old}(m|x_n)$ is computed as matrix $P \in \mathcal{M}(N_d \times N_r)$ with entries

$$p_{mn} = \frac{\exp\left(\frac{\|x_n+V_{x_n}d_n-(R(y_m+V_{y_m}d_m)+t)\|^2}{-2\sigma^2}\right)}{\sum_{k=1}^{N_d}\exp\left(\frac{\left\|x_n+V_{x_n}d_n-\left(R\left(y_k+V_{y_k}d_k\right)+t\right)\right\|^2}{-2\sigma^2}\right)} . \quad (6)$$

Eq. (5) evaluates to

$$Q_{rigid}(R,t) = \\ \sum_{n=1}^{N_r}\sum_{m=1}^{N_d} P^{old}(m|x_n)\frac{\|x_n+V_{x_n}d_n-(R(y_m+V_{y_m}d_m)+t)\|^2}{2\sigma^2} . \quad (7)$$

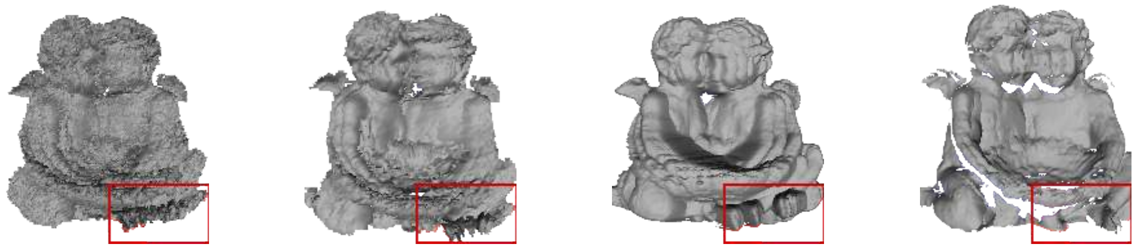
In the case II, the same expression reads

$$Q_{non-rigid}(d_1,\dots,d_O) = \\ \sum_{n=1}^{N_r}\sum_{m=1}^{N_d} P^{old}(m|x_n)\frac{\|x_n+V_{x_n}d_n-(R(y_m+V_{y_m}d_m)+t)\|^2}{2\sigma^2} \\ +\lambda\left\|(d_1-d_2,\dots,d_{O-1}-d_O)\right\|^2 . \quad (8)$$

The format of the entries of $P$ follows accordingly. The complete EM procedure for aligning all $Y_\ell$ against the reference set $X$ is given in Figure 5 as pseudocode. The optimizer terminates if there is no further improvement or the maximum number of iterations has been reached. The result is the set of rigid alignment parameters $(R_\ell, t_\ell)$, $\ell = 1,\dots,K$, as well as the systematic bias values $d_1,\dots,d_O$. We remark that the bias values are assumed to be the same for all high resolution scans.

Please note that for efficiency reasons, in case II we don't evaluate $Q_{non-rigid}$ for all 3D points, but only for a subset of samples from $X$ and the $Y_\ell$. Fig. 4b illustrates this sampling pattern in the pixel (depth image) domain. The red lines are the depth pixels (3D points) of $X$ which are included, and the green elliptical regions are the ones from each $Y_\ell$ included. Also, for camera paths covering a larger viewpoint range, we perform several global alignments to several reference scans, such that sufficient overlap is guaranteed.

Due to space limitations we do not provide all mathematical expressions in detail, but provide an additional document with more specifics.



(a) ICP with 70 Raw Inputs

(b) ICP with 7 Averaged Raw Inputs

(c) ICP with 7 Superresolved Inputs

(d) Proposed Method

Figure 6: All steps in our pipeline are important. (a) Aligning 7 chunks of frames (10 frames each) with ICP fails (severe noise, wrong alignment - *e.g.* around feet (red)). (b) averaging 10 frames from the center view of each chunk center reduces noise, but does not provide more resolution. ICP alignment fails. (c) Superresolution of the chunks boosts resolution, but ICP fails due to non-rigid distortions. (d) our method produces clear detail and correct alignment. (Note: no hole-filling was done; triangles at occlusion edges were filtered).

## 4. Results

We have tested our approach with four different test objects: an antique head (height $31$ cm), a vase (height $47$ cm, a Buddha statue (height $59$ cm), and a sculpture of two angels (height $34$ cm), Fig. 1 and 8 and supplemental video. For each object we also captured a ground truth model with a Minolta Vivid 3D scanner. We have chosen these objects due to their reasonable size and surface detail below the ToF resolution. Of each object we captured $600$ frames on an arc covering around $120°$ around the object. The distance of the camera to the object was about $140$ cm in all cases. We used two different setups to capture the object. In the first setting the object was placed on a turn table that was manually turned in front of the static camera (exposure time 30 ms). For all figures but the angels in Fig. 8k-8o this setup was used. The angels were scanned with a hand-held setup where the camera was manually moved moved around the object. Here the integration time was 35 ms. We plan to make the data sets and results available for download.

For all scenes we used $K = 7$ chunks with frames $5, 105, \cdots, 505, 595$ being the center frames for the chunks. The chunk size was always $C = 10$ frames. All of the 3D renderings of our results in this paper were created by converting the individual 3D scans (low-resolution and high-resolution) to meshes and render them in the same coordinate system. Meshes are created by regular triangulation according to the depth pixel order in each depth map. Triangles that span occlusion boundaries were filtered out as good as possible, but some remaining jaggy edges may remain in parts. Note that this is not a reconstruction artifact but a side-effect of our rendering method. We did no hole filling, neither on the laser scans nor on our models, in order to display the original output of our method. We also abstain from explicit scan merging or single surface reconstruction as this is not the main focus of this paper and may lead to unwanted smoothing.

In a single ToF scan (Fig. 2 and Fig. 8) one can hardly even recognize the overall type of an object, let alone fine detail. Please note that in some raw depth scans depth pixels with low return amplitude have been discarded, as they are considered unreliable by the superresolution approach, see Sect. 3.1. The superresolution approach creates high-resolution denoised depth scans for certain viewpoints (the chunks), as shown in Fig. 2. By this means, the resolution and visible detail in each such scan is already dramatically improved, but each map only covers a part of the object and suffers from nonlinear distortion due to the camera bias. Our final alignment registers the scans into a 3D model of reasonable quality, Fig. 8 and 1. A lot of finer scale detail that lies close to the resolution limit of the ToF camera comes out in the final results, such as the petals of the flower in the angels sculpture, or the folds in the robe of the Buddha statue. Note that the holes in some parts, *e.g.* in the flower, are due to the camera path where this region was occluded. As stated above, we purposefully refrain from hole-filling.

Of course, the resolution and shape quality of our models cannot rival that obtainable with a laser scanner, Fig. 8 and 1, and we never claimed that. Nonetheless, our models are of sufficient quality for many applications were submillimeter accuracy is not required, a result that was at first unexpected looking at the raw ToF quality.

**Validation** Each step in our algorithm is important for the success of the method. To demonstrate this, we show in Fig. 6 the failures of a few alternative strategies which one may have considered: First, aligning raw depth scans using ICP fails due to severe noise and non-rigid bias-induced distortions of the individual 3D scans, Fig. 6a. Averaging 10 ToF scans at 7 fixed viewpoints around the object reduces random noise, but does not boost resolution. Alignment with ICP still fails, Fig. 6b. Performing 3D superresolution for 7 chunks around the object and aligning with ICP boost detail, but still fails to align the scans properly as visible by the multiple layers of geometry in the arms, Fig. 6c. This is due to the fact that the systematic bias also non-rigidly distorts the superresolved geometry. In contrast, the combination of ToF 3D superresolution and our new probabilistic scan alignment produces correctly aligned results with clearly visible detail, Fig. 6d.

We also quantitatively measured the reconstruction quality against the laser scanned ground truth. As one can see in the color-coded reconstruction error renderings in Fig. 8 and 1 our models compare very favorably. In most areas the error is below $1.0$ cm, and there are only a few outliers. In the figures, areas where there is no ground truth geometry in the laser scan (*e.g.* holes due to occlusion) are rendered in grey.

Fig. 7 plots the bias distribution against the pixel distance from the depth map center for all four objects. The bias distribution is fairly stable across objects, and follows our assumption of monotonous increase with radius. Slight differences between the four scenes exist. This can be due to slight scene dependencies which our algorithm can accommodate for in certain bounds.

Throughout our experiment we used the same parameters for the LidarBoost ($\lambda_{SR} = 1, 5 \times 5$ regularizer region, see [23]) algorithm and our alignment procedure ($\lambda = 3$). This shows that our method is rather stable and does not require scene dependent parameter tuning. With an unoptimized MATLAB implementation it takes around 330 min to run the entire pipeline for one data set.

**Limitations** Our approach is subject to a few limitations: the ToF camera fails to capture good data for certain surface materials, like highly specular objects. However, other scanners suffer from similar limitations. We use a structure from motion approach for initialization, so certain features must be visible in the amplitude images. Other non-optical initialization would be feasible. While scan acquisition is straightforward and fast, the runtime of scan reconstruction is notable. However, almost $95\%$ of the current runtime is due to the superresolution. We expect dramatic speedups by using a gradient-based optimizer rather than the one suggested by [23] which we currently use. Despite these limitation, we were able to demonstrate that good quality 3D shape scanning without manual correction is also feasible with low quality sensors.

## 5. Conclusion

In this paper we demonstrated that 3D shape models of static objects can also be acquired with a Time-of-Flight sensor that, at first glance, seems completely inappropriate for the task. The key in making this possible is the effective combination of 3D superresolution with a new probabilistic multi-scan alignment algorithm tailored to ToF cameras.

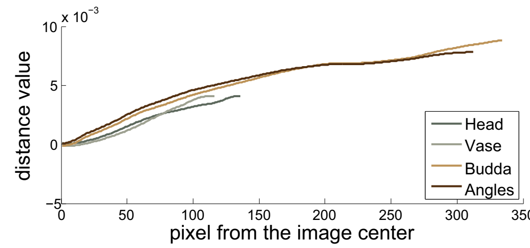



Figure 7: Bias factors with increasing pixel distance from the center for the four test objects. Note that the vase and the head cover less field of view in the ToF sensor, thus the curves terminate earlier.

In future, we plan to investigate approaches for real-time shape scanning, as well as incorporation of more sophisticated noise models into the reconstruction framework.

## References

[1] D. Anderson, H. Herman, and A. Kelly. Experimental characterization of commercial flash ladar devices. In *Proc. of ICST*, 2005.

[2] C. Beder, B. Bartczak, and R. Koch. A combined approach for estimating patchlets from pmd depth images and stereo intensity images. In *Proc. DAGM*, pages 11–20, 2007.

[3] R. Benjemaa and F. Schmitt. A solution for the registration of multiple 3d point sets using unit quaternions. In *Proc. ECCV '98 II*, pages 34–50, 1998.

[4] R. Bergevin, M. Soucy, H. Gagnon, and D. Laurendeau. Towards a general multi-view registration technique. *IEEE PAMI*, 18(5):540–547, 1996.

[5] P. J. Besl and N. D. McKay. A method for registration of 3-d shapes. *IEEE PAMI.*, 14(2):239–256, 1992.

[6] J.-Y. Bouguet and P. Perona. 3d photography on your desk. In *Proc. ICCV*, page 43. IEEE, 1998.

[7] B. Brown and S. Rusinkiewicz. Global non-rigid alignment of 3-D scans. *ACM (Proc. SIGGRAPH)*, 26(3), Aug. 2007.

[8] B. Curless and M. Levoy. A volumetric method for building complex models from range images. In *Proc. SIGGRAPH*, pages 303–312. ACM, 1996.

[9] J. R. Diebel, S. Thrun, and M. Brünig. A bayesian method for probable surface reconstruction and decimation. *ACM TOG*, 25(1):39–59, 2006.

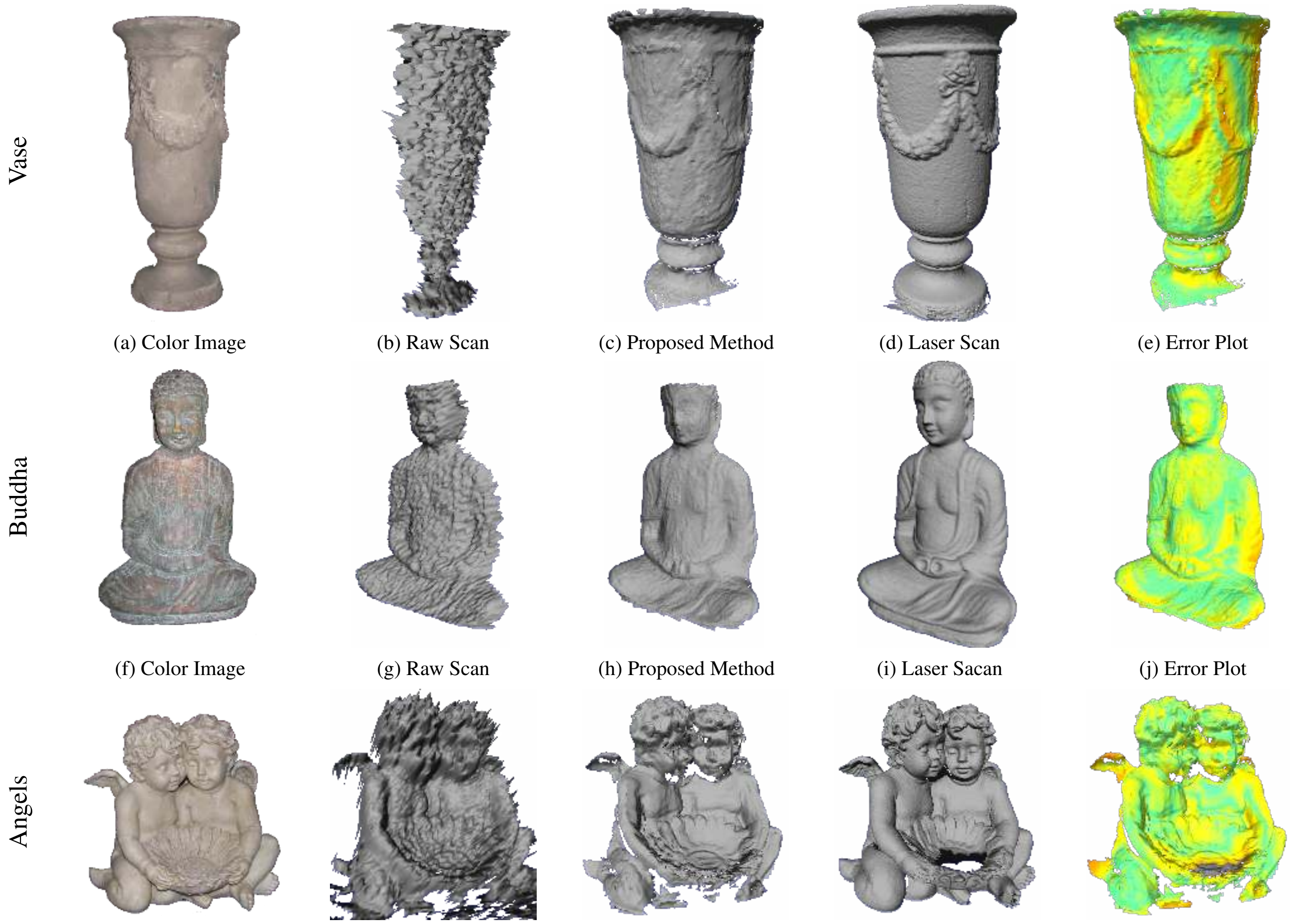
[10] Y. Furukawa and J. Ponce. Accurate, dense, and robust multi-view stereopsis. In *CVPR '07*, 2007.

[11] Q.-X. Huang, S. Flöry, N. Gelfand, M. Hofer, and H. Pottmann. Reassembling fractured objects by geometric matching. *ACM TOG.*, 25(3):569–578, 2006.

[12] B. Huhle, P. Jenke, and W. Straßer. On-the-fly scene acquisition with a handy multi-sensor system. *International Journal of Intelligent Systems Technologies and Applications*, 5(3):255–263, 2008.

[13] P. Jenke, M. Wand, M. Bokeloh, A. Schilling, and W. Straer. Bayesian point cloud reconstruction. *Computer Graphics Forum*, 25(3):379–388, September 2006.

[14] Y. J. Kil, B. Medereos, and N. Amenta. Laser scanner super-resolution. In *Point-based Graphics*, 2006.




(a) Color Image (b) Raw Scan (c) Proposed Method (d) Laser Scan (e) Error Plot

(f) Color Image (g) Raw Scan (h) Proposed Method (i) Laser Sacan (j) Error Plot

(k) Color Image (l) Raw Scan (m) Proposed Method (n) Laser Scan (o) Error Plot

Figure 8: Results of our method - in each row: test object; single ToF depth scan; our reconstructed model; laser scanned model; color-coded distance error against laser scan. Our algorithm reconstructs 3D models of reasonable quality despite severely distorted raw ToF data. (Note: raw aligned scans, no hole filling done)

[15] Y. Kim, D. Chan, C. Theobalt, and S. Thrun. Design and calibration of a multi-view tof sensor fusion system. In *Proc. CVPR Worksh. TOF-CV*, pages 1–7, 2008.

[16] A. Kolb, E. Barth, R. Koch, and R. Larsen. Time-of-Flight Sensors in Computer Graphics. *EUROGRAPHICS STAR report*, 2009.

[17] R. Lange. 3D time-of-flight distance measurement with custom solid-state image sensors in CMOS/CCD-technology. *Diss., University of Siegen*, 2000.

[18] D. Lanman and G. Taubin. Build your own 3d scanner: 3d photograhy for beginners. In *SIGGRAPH courses*, pages 1–87. ACM, 2009.

[19] M. Lindner and A. Kolb. Lateral and depth calibration of PMD-distance sensors. *Advances in Visual Computing*, pages 524–533, 2006.

[20] A. Myronenko, X. Song, and M. Carreira-Perpinan. Non-rigid point set registration: Coherent Point Drift. *NIPS*, 19:1009, 2007.

[21] A. Rajagopalan, A. Bhavsar, F. Wallhoff, and G. Rigoll. Resolution Enhancement of PMD Range Maps. *LNCS*, 5096:304–313, 2008.

[22] S. Rusinkiewicz, O. Hall-Holt, and M. Levoy. Real-time 3d model acquisition. In *Proc. SIGGRAPH*, pages 438–446. ACM, 2002.

[23] S. Schuon, C. Theobalt, J. Davis, and S. Thrun. Lidarboost: Depth superresolution for tof 3d shape scanning. *Proc. CVPR*, 2009.

[24] S. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski. A comparison and evaluation of multi-view stereo reconstruction algorithms. In *Proc. CVPR*, volume 1, 2006.

[25] A. Swadzba, B. Liu, J. Penne, O. Jesorsky, and R. Kompe. A comprehensive system for 3D modeling from range images acquired from a 3D ToF sensor. In *Proc. ICCV*, 2007.

[26] J. Zhu, L. Wang, R. Yang, and J. Davis. Fusion of time-of-flight depth and stereo for high accuracy depth maps. *Proc. CVPR*, 0:1–8, 2008.