Towards Precise Real-Time 3D Difference Detection for Industrial Applications

Svenja Kahn[a], Ulrich Bockholt[a], Arjan Kuijper[a,b], Dieter W. Fellner[a,b]

[a] *Fraunhofer IGD, 64283 Darmstadt, Germany*
[b] *Department of Computer Science, TU Darmstadt, 64289 Darmstadt, Germany*

Abstract

3D difference detection is the task to verify whether the 3D geometry of a real object exactly corresponds to a 3D model of this object. We present an approach for 3D difference detection with a hand-held depth camera. In contrast to previous approaches, with the presented approach geometric differences can be detected in real-time and from arbitrary viewpoints. The 3D difference detection accuracy is improved by two approaches: First, the precision of the depth camera's pose estimation is improved by coupling the depth camera with a high precision industrial measurement arm. Second, the influence of the depth measurement noise is reduced by integrating a 3D surface reconstruction algorithm. The effects of both enhancements are quantified by a ground-truth based quantitative evaluation, both for a time-of-flight (SwissRanger 4000) and a structured light depth camera (Kinect). With the proposed enhancements, differences of few millimeters can be detected from one meter measurement distance.

Keywords:
3D imaging, depth cameras, geometric difference detection, 3D discrepancy check

1. Introduction

Depth cameras have evolved rapidly in the last years. They are able to acquire dense surface measurements at an framerate of up to 30 frames per second. The most promising depth camera technologies are time-of-flight depth cameras [1] [2] and structured light depth cameras (which calculate the depth from patterns projected onto the scene), such as the Kinect depth camera [3]. Whereas depth cameras are commonly used in the consumer mass market (for example in gaming applications), up to now they are only seldomly used in industrial applications. This is mainly due to the poor measurement quality of these depth cameras: Depending on the surface properties of the measured object, the measured distance can differ from the real distance by several centimeters at a distance from 0.5 to 5 meters.

For industrial applications which require very precise 3D measurements, high-end laser scanners are the technology of choice [4] [5]. While 3D imaging based on stereo vision requires a texture on the scanned objects, laser scanners can accurately measure the shape of untextured objects. In contrast to other passive 3D distance estimation approaches such as shape from shadow or shape from shading [6], laser scanners do not require previous knowledge about the surface reflectance properties of the scanned objects or about the lighting of the scene. Just as laser scanners, state of the art depth cameras (based on variants of structured light or time-of-flight based depth imaging) neither require textures nor previous knowledge about the surface properties. While laser scanners offer a very accurate measurement precision (up to sub-millimeter accuracy), they are very expensive and not always eye-safe. Furthermore, they usually only capture depth measurements along a single scan line per time instance, instead of a full depth image as acquired by depth cameras. To acquire a complete 3D point cloud, these point- or line based laser scanners need to sequentially scan the environment, either by automatically rotating parts of the scan head [7][8] or with hand-held approaches [9]. Each scan takes several seconds to several minutes. In contrast, depth cameras acquire dense 3D point clouds at interactive update rates of up to 30 frames per second.

One application area which is strongly limited by the drawbacks of custom laser scanning is 3D difference detection. This is the task to verify whether the 3D geometry of a real object exactly corresponds to a 3D model of this object, or whether there are differences between the 3D model and the real object. This is of particular importance for industrial tasks such as prototyping, manufacturing and assembly control. Despite their limited measurement accuracy, the application of depth cameras for 3D difference detection has a lot of potential: Depth cameras are low-cost devices which are able to capture the 3D surface of an object in real time, which can be moved during data acquisition and which are eye-safe for the users. In contrast to stationary laser scanners, the user can move the depth camera around the object. Thus, the user is not restricted to a single viewpoint, but can inspect differences at arbitrary parts of the object from arbitrary viewpoints in real time.

An application example for real-time geometric 3D difference detection based on dense real-time 3D imaging with depth cameras is assembly control: After a worker has assembled several parts of an object, geometric difference detection between a reference 3D model and the assembled object can be used to check if each component was attached at the correct position. The same approach can also be used to immediately detect differences during the assembly process itself. Such a discrepancy check can for example be used to detect if a tube or a pipe was attached to a different position than intended. Another application example is manufacturing and prototyping. Given a 3D model of the manufactured object or the prototype, a 3D discrepancy check can detect differences between the 3D model and the constructed object which might occur due to inaccuracies in the manufacturing process or due to changes during the prototyping process. 3D difference detection for construction is a third application area. After a building element or a technical installation was constructed, 3D difference detection can be used to check whether the constructed and installed elements really comply to the 3D specification.

The most important step for making the use of depth cameras feasible for industrial 3D difference detection is the algorithmic integration of methods which enhance the accuracy of depth image based 3D difference detection, both in terms of the accuracy of the camera pose estimation and in terms of an enhancement of the measurement quality. The work presented in this paper was in parts developed in a cooperation between Fraunhofer and Volkswagen research, with the goal to make the use of depth image based 3D difference detection feasible for industrial ap-

plications. In this paper, we first explain how depth cameras can be used to detect differences between an object and a 3D model of this object and describe the major sources for inaccuracies in the 3D difference detection process (Section 2). In Section 3, we describe different methods for estimating the camera pose and present our approach to enhance 3D difference detection with a high precision pose estimation by a mechanical measurement arm. Section 4 gives an overview of approaches for reducing depth measurement inaccuracies and describes the integration of a 3D reconstruction algorithm [10][11] into the 3D difference detection pipeline. This on-the-fly 3D reconstruction of the captured object surface reduces measurement inaccuracies by fusing depth measurements from several depth images. Section 5 provides a quantitative evaluation based on real 3D measurements with known ground truth data. This evaluation quantifies the achievable accuracy of depth image based 3D differences detection, both for a Swissranger 4000 time-of-flight depth camera and a structured light depth camera (Kinect). Furthermore, this evaluation quantifies the accuracy enhancement of the 3D difference detection by the integration of the proposed accuracy enhancement methods (high precision pose estimation with a measurement arm and measurement enhancement by fusing several depth measurements). Finally, the evaluation results are discussed in Section 6.

2. Geometric 3D Difference Detection

In the last years, first approaches for geometric 3D difference detection have been developed which capture the real objects either with 2D cameras or with 3D depth cameras. After a brief overview of the state-of-the art, this section describes the error sources which limit the accuracy of depth image based geometric difference detection and which need to be overcome to make the use of depth image based difference detection feasible for industrial applications.

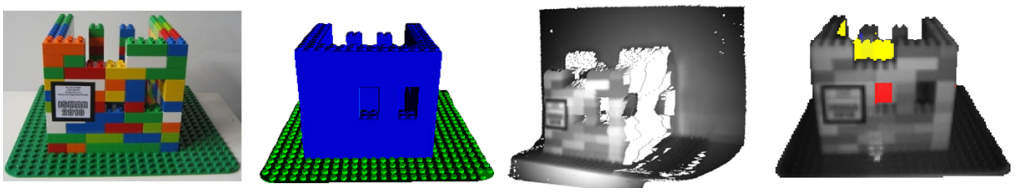
2.1. 2D Camera based Difference Detection

The first solutions for camera based difference detection were based on still 2D images captured with 2D color cameras: In their pioneering work Georgel et al. presented an augmented reality solution for discrepancy check in the context of construction planning [12] [13] [14]. Their system allows engineers to superimpose 2D photographs of a plant with the CAD model developed during the planning phase. Whereas this augmentation of still 2D images with the 3D model is very useful to visually compare the 3D model and the real scene, it is limited to the 2D information contained in the images and provides no possibility to automatically compare the 3D data of the model with the 3D geometry of the real scene. Webel et al. presented a system for AR discrepancy check with which the 3D positions of single points in the 3D model and the real scene can be compared [15]: A laser pointer is used to depict a point on the surface of the real scene. The 3D coordinate of the point is reconstructed by triangulation with a stereo camera system. Whereas this approach allows the comparison of single 3D points, it is not suited for dense 3D difference detection.

2.2. 3D Difference Detection with Depth Cameras

Kahn et al. presented an approach for real-time geometric 3D difference detection with a hand-held depth camera [16] [17]. The depth camera is used to capture the distances to the surface of an object. Then, the differences between these measurements and a 3D model of the captured object are estimated for each acquired 3D measurement. Previous approaches using 3D

measurements were either restricted to a static camera position [18] or required the manual specification of 3D correspondences between a laser scan of a construction site and a 3D model of the construction site for each new scan position [19] [20], to transform both data sets into a common coordinate system. In contrast, the approach proposed for depth camera based difference detection is feasible for interactive real-time 3D difference detection with a moving camera.



(a) Real object (b) Virtual 3D model (c) 3D point cloud (d) Visualized 3D differences

Figure 1: Depth image based 3D difference detection from arbitrary camera positions (early demonstrator).

Figure 1 shows an early, basic demonstrator for real-time geometric difference detection with a hand-held depth camera. The coordinates of the square marker attached to the brick model (Figure 1a) are specified in the coordinate system of the virtual 3D model (Figure 1b). By detecting the image marker in the 2D camera image and estimating the camera pose with image-based camera tracking, the position and orientation of the depth camera is calculated relative to the 3D model's coordinate system. Then, the 3D model is rendered from the estimated pose of the depth camera to acquire an artifical depth image from the depth buffer of the graphics card. Each such synthesized depth value is compared to the real depth measurement acquired by the depth camera. Figure 1c shows a rotated view of the 3D point cloud acquired with a time-of-flight depth camera. In Figure 1d, the detected differences are visualized with a color based augmentation of the depth camera's intensity image.

2.3. From First Approach to Improved Accuracy: Factors Which Cause Inaccuracies

While the difference detection of the previously described basic approach provides the basic aspects for real-time depth image based 3D difference detection, it is far too inaccurate for being applicable for industrial scenarios. With this basic approach, only very distinct differences can be detected. The plate of the 3D model shown in Figure 1 has a size of 38×38 cm. The distance of the time-of-flight depth camera to the 3D model is about 50-90cm and only differences bigger than 5cm are visualized. Reducing this difference visualization threshold could reveal smaller differences. However, this would also cause the visualization of seeming differences which are caused by inaccuracies in the difference detection process.

To make the use of depth cameras feasible for real-time geometric difference detection in industrial applications, it is necessary to identify the factors which cause inaccuracies in the geometric difference detection process and to develop solutions which reduce these inaccuracies. The causes for inaccuracies in the 3D difference detection process can be classified in two categories:

1. Inaccuracies in the depth camera's pose estimation
    - Inaccuracies of the pose estimation device

- Inaccurate relative transformation between depth camera and pose estimation device
- Inaccurate alignment of world coordinate system and 3D model coordinate system
- Temporal offset between pose acquisition by pose estimation device and depth image acquisition by depth camera

2. Measurement inaccuracies of the depth camera
   - Random measurement noise
   - Systematic measurement errors
   - Motion blur effects (reducing the accuracy of 3D measurements)

Inaccuracies in the Depth Camera's Pose Estimation. The pose of the depth camera can either be estimated based on the data captured by the depth camera itself, or by an additional pose estimation device (such as an additional, higher resolution 2D color camera used for the image based camera tracking [16] [17]). Either way, inaccuracies in the pose estimation reduce the accuracy of the difference detection. If an additional pose estimation device is used, the overall accuracy also depends on the accuracy of the estimated relative transformation between the depth camera and the pose estimation device. Furthermore, the accuracy decreases if the world coordinate system (in which the pose of the camera is estimated) is not correctly aligned with the 3D model coordinate system (for example, if the 3D coordinates of the marker used for the camera pose estimation are not very accurate). Finally, the accuracy also decreases if the depth image was acquired at another timestamp than the time at which the pose was estimated. Section 3 provides an overview of approaches which can be used to enhance depth image based 3D difference detection by improving the pose estimation accuracy.

Measurement Inaccuracies of the Depth Camera. The depth measurements of depth cameras are affected by random noise as well as by systematic errors. The random noise causes the effect that even for a static scene and a static position of a depth camera, the measured distances differ from frame to frame. For state-of-the art depth cameras, the random measurement noise is typically stated to be about 1% of the distance to the camera [21]. The measurement noise can be modelled with a Gaussian distribution. In contrast to the measurement noise, the systematic errors are not characterized by varying distance measurements for a constant distance, but by a systematic offset. For example, systematic errors occur due to reflections of specular surfaces or due to a constant offset between the measured and the real differences in dependance of the distance of the measured object to the camera center. The systematic errors are specific for each different depth measuring technology and were evaluated in previous publications, both for time-of-flight depth cameras [22] [23] [24] [25] and for the Kinect depth camera [26] [27]. The Kinect depth sensing technology was developed by PrimeSense [28]. This depth sensing technology is also used in other recent depth cameras, such as the Asus Xtion depth camera. Different depth cameras based on the PrimeSense technology have very similar measurement properties [29]. Current state-of-the art depth cameras are usually either based on this technology, or on the time-of-flight depth measurement principle. Section 4 provides an overview of approaches for enhancing depth image based 3D difference detection by improving the accuracy of acquired depth measurements.

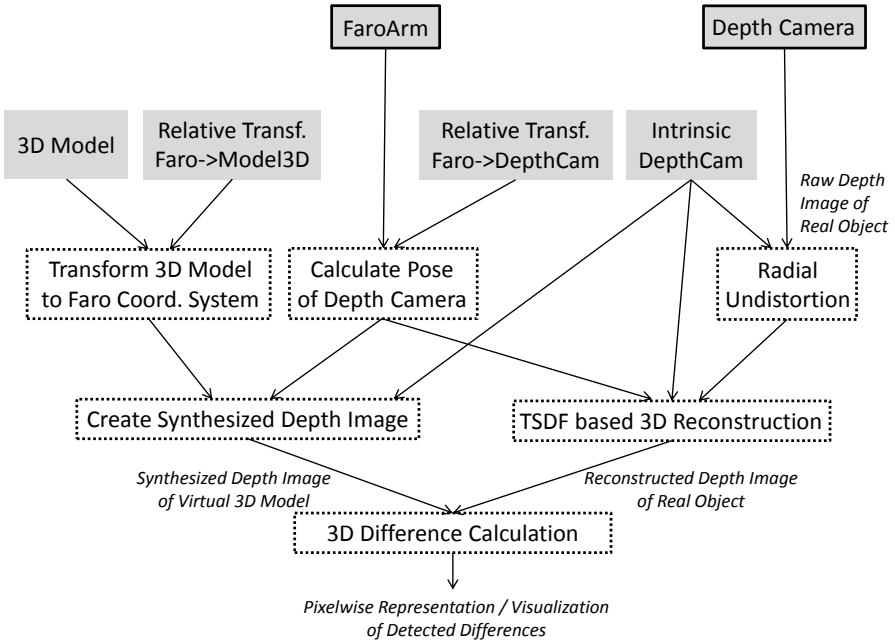


Figure 2: 3D difference detection pipeline.

Enhanced 3D Difference Detection Pipeline. Figure 2 provides an overview of the 3D difference detection pipeline. This data flow diagram already contains the two approaches for enhancing the accuracy of the 3D difference detection which are explained in the following two chapters: high precision pose estimation with a Faro arm and accuracy enhancement of the 3D measurements by a 3D surface reconstruction algorithm. In the basic approach for which results were shown in Figure 1, the camera pose was estimated with a marker tracker. Furthermore, the 3D reconstruction was not yet integrated into the pipeline.

3. Enhancing 3D Difference Detection with an Accurate Pose Estimation and Registration

For comparing a depth camera's 3D measurements of an object with a 3D model of this object, a correspondency problem needs to be solved: Which 3D measurement in the depth image corresponds to which 3D position in the 3D model? For a depth camera, this task is equivalent to the estimation of the position and orientation of the depth camera relative to the 3D model. This section describes and discusses three possible approaches to solve this task: Image based camera tracking, geometric registration of the depth images and pose estimation with a high precision mechanical measurement arm.

3.1. Image based Camera Pose Estimation

Previous approaches for 3D difference detection used image based camera tracking for the pose estimation [16][17] . With image based camera tracking, the camera pose can coarsely be estimated from the intensity image of a time-of-flight depth camera. Another approach is

to rigidly couple the depth camera with a higher resolution 2D color camera whose 2D image is used for the camera tracking. In both cases, the camera pose can be estimated with a combination of marker-based camera tracking and structure from motion 3D reconstruction [30]. While the camera is moved, this approach detects and tracks characteristic image features [31][32]. Then the 3D positions of the tracked features are reconstructed online via triangulation. Finally the 2D-3D correspondences between the 2D image coordinates of the tracked features and their reconstructed 3D coordinates are used to calculate the camera pose if no marker is visible.

Image based camera pose estimation has the advantage that it can be computed efficiently and that it is real-time capable. However, the accuracy of image based camera tracking strongly depends on the captured scene: The camera pose can only be estimated with an image based approach if enough stable, characteristic image features are visible in the captured camera image. Thus, image based camera pose estimation cannot be used to estimate the camera pose in environments that contain large surface parts which are uniformly colored and which are not well textured. This is often the case in industrial scenarios, which might contain many untextured or metallic materials which are difficult to track with image based camera pose estimation. Furthermore, the accuracy of image based camera tracking is sensitive to the distribution of the detected features in the 2D camera image. The accuracy strongly decreases if the tracked features are not evenly distributed in the whole camera image (for example, if features can only be found in a part of the camera image). With image based camera pose estimation, the estimated camera pose can differ from the real camera pose by several millimeters to centimeters, which strongly affects the accuracy of the overall 3D difference detection approach.

3.2. Geometric 3D Registration

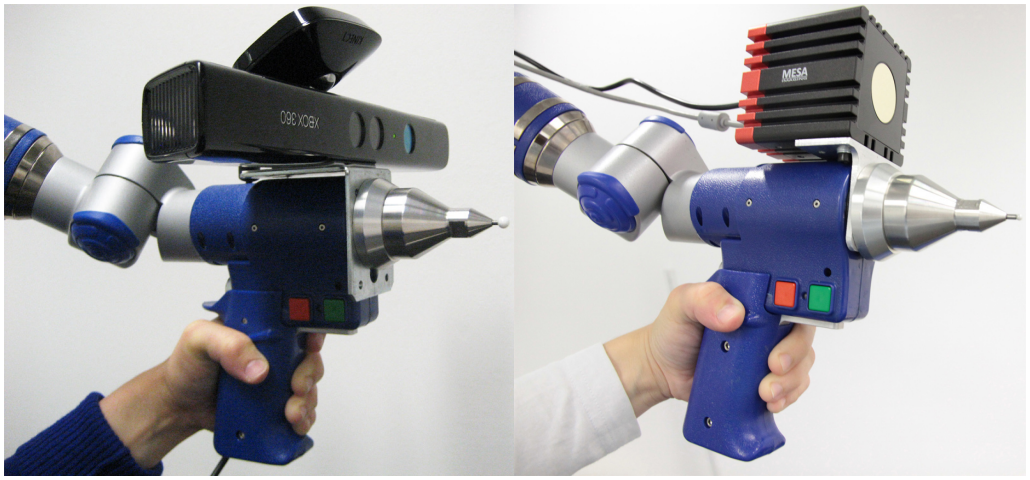
As an alternative to 2D image based camera tracking, the depth measurements acquired by the depth camera can also be used to calculate the pose of the depth camera. The pose can either be estimated by geometrically aligning the current depth image with the provided 3D model or by aligning the current depth image with 3D measurements acquired from previous depth images. The Iterative Closest Point algorithm [33] is a geometric registration algorithm which can be used to align the current depth image with previously captured 3D data. This algorithm iteratively estimates a transformation which minimizes the 3D distances between the depth image and the 3D model or the previously reconstructed 3D point cloud. Geometric registration is computationally very expensive (especially, if a large number of 3D-3D correspondences are aligned to find a robust solution for the registration). Without massive subsampling, geometric registration on the CPU is not real-time capable for the large number of 3D points captured by a depth camera. In contrast to CPU based geometric registration, Newcombe et al. [10] showed that a geometric real-time registration is feasible with a highly parallelized implementation on a modern graphics card. Such a GPU-based registration could either be used as a standalone method to estimate the pose of the depth camera, or it could be used to improve the accuracy of a pose estimated with image based camera tracking. However, similar to image based camera tracking, the achievable accuracy of camera pose estimation based on geometric features strongly depends on the structure of the captured scene.

Geometric camera pose estimation requires the captured scene to be structured in such a way that the camera pose can be calculated unambiguously from the depth image. This condition is often not fulfilled. For example, if a depth camera captures several parallel pipes in front of a wall, the camera pose cannot be calculated unambiguously from the depth image as there is one degree of freedom along the pipes. The accuracy of geometric camera pose estimation also depends on the movement path of the camera. For example, the camera pose often cannot be

estimated robustly when the user moves the camera close to the surface to inspect details, or when the user increases the distance between the camera and the object surface. Furthermore, it is often not obvious whether the estimated camera pose and the geometric alignment have drifted and need to be reset, and if so, how the pose estimation and the alignment can be robustly re-initialized while keeping previously captured data.

3.3. Pose Estimation with a Precise Measurement Arm

Due to the drawbacks of both image based and geometric alignment based camera pose estimation, we propose to use an industrial measurement arm for the pose estimation of the depth camera. Such a mechanical measurement arm outputs a very precise pose estimation for its measuring tip with a guaranteed precision better than 0.1mm. By rigidly coupling a depth camera with the measuring tip of such a measurement arm, the pose of the depth camera can also be estimated with a high precision. Figure 3 shows two depth cameras rigidly coupled with a high precision Faro Platinum measurement arm. In contrast to both image based and geometric alignment based pose estimation, the accuracy of the pose estimation is completely independent of the captured scene. With a measurement arm, the camera pose estimation cannot fail due to ambiguities or due to too few optical or geometric features. Thus, the user does not need to be careful when moving the camera to avoid tracking loss. Instead, the user can fully concentrate on the 3D difference detection task at hand. Figure 4 shows how a depth camera rigidly coupled with a measurement arm can be used to detect differences between a real mockup and a 3D model of this object.



(a) Kinect (b) SwissRanger 4000

Figure 3: Depth cameras coupled with a Faro measurement arm [34].

Pose estimation with a measurement arm has the advantage that only minimal computational ressources are required for estimating the camera pose: The measurement arm internally estimates the pose of the measurement tip by the rotational angles of its rotational joints and outputs the pose of the point tip. Given the pose of the measurement tip, only the positional and rotational offset between the depth camera and the measurement tip needs to be added to infer the

pose of the depth camera from the pose of the measurement arm. This offset is calculated with an offline hand-eye calibration [35] [36], which only needs to be calculated once for each depth camera. For a depth camera, this relative transformation between the tip of the measurement arm and the depth camera can be estimated in a similar way as the hand-eye calibration between a robot (or a measurement arm) and a 2D color camera. Therefore, the 3D coordinates of a 2D calibration pattern (for example a checkerboard or a well detectable marker) are measured with the tip of the measurement arm. Then, the calibration pattern is detected in the 2D image of the depth camera. Time-of-flight depth cameras directly measure an intensity (grey) value for each captured pixel, which can be used to detect the marker. In contrast to time-of-flight cameras, per default the depth camera of the Kinect only outputs a depth value per pixel and no color or intensity information. However, for example with the OpenNI interface [37], the Kinect depth camera can be switched from depth to infrared acquisition mode. Thus, it is possible to detect the calibration pattern in the infrared image of the depth camera and to estimate the pose of the Kinect depth camera based on the detected 2D pattern. Finally, the hand-eye transformation is calculated from a large set of pose pairs where each pair stores the pose of the measurement arm and the estimated pose of the depth camera (see [34] for details).

If the hand-eye calibration between the depth camera and the measurement arm was calculated with a high precision, the pose of a depth camera rigidly attached to a measurement arm can be expected to be rather precise because the other causes for pose inaccuracies listed in Section 2.3 are handled well by a measurement arm: First, the guaranteed precision of a Faro measurement arm is better than 0.1 mm and thus much more precise than the accuracy of image based camera tracking. Second, the surface of the real object can be measured very precisely with the point tip of the measurement arm. Thus, the relative transformation between the real object, the 3D model and the coordinate system of the measurement arm can be estimated with a high precision. Third, using such an industrial measurement arm for the pose estimation also has the advantage that it estimates the pose with a high update rate (>60fps). In comparison to pose estimation with an unsynchronized color camera (<30fps), this minimizes the temporal offset between the pose estimation and the acquisition of the depth images and thus contributes to an enhancement of the overall accuracy.

4. Enhancing 3D Difference Detection by Reducing Measurement Noise

The accuracy of a depth camera based 3D difference detection strongly depends on the accuracy of the acquired depth values. However, for current state-of-the art depth cameras, the measured depth values can differ from the real distance by several millimeters to centimeters. As described in Section 2.3, the measurement accuracy is decreased both by systematic measurement errors and by random measurement noise. Several approaches have been proposed to reduce the systematic measurement errors of specific depth cameras, both for different time-of-flight depth cameras [22] [23] [24] [25] and for the Kinect depth camera [26] [27].

In contrast to previous approaches (which investigated depth camera calibration procedures to reduce the systematic measurement errors), this section focuses on the reduction of the non-systematic, random measurement noise. For static camera positions, the measurement noise can be reduced easily by capturing several depth images from the same static camera position and by averaging all depth values acquired at a pixel over time. However, this is not possible for a moving depth camera because depth values acquired at a certain pixel correspond to different 3D points of the captured scene.

Figure 4: 3D difference detection with a Kinect depth camera and a Faro measurement arm.

For this reason, we conducted a literature research of approaches which can be used to reduce the measurement noise for moving depth cameras. Most of these approaches are superresolution methods, which either combine the depth image with information from an additional color image or which fuse several depth images taken from close camera positions. However, recently another approach (KinectFusion [10]) has been proposed. In contrast to superresolution methods, this approach does not reconstruct a single improved depth image but reconstructs an implicit 3D surface representation of the captured scene. The remainder of this section describes both superresolution and the surface reconstruction approach and discusses the applicability of these approaches for 3D difference detection with a hand-held depth camera.

4.1. Superresolution

Superresolution methods improve the resolution and the quality of images, for example by combining several images taken from very close camera positions. While early superresolution methods were targeted at 2D images [38], recently also superresolution methods have been developed for depth images. Depth image based superresolution approaches can be divided into two categories: First, methods which combine a depth image with a higher resolution 2D image. Second, methods which fuse the depth data of several depth images.

Superresolution approaches which use a higher resolution 2D image for enhancing the depth values of a depth image [39] [40] [41] are based on the assumption that the depth values correlate with the color values of the 2D images. This assumption is met if depth edges in the depth image correspond to color edges in the 2D image and if uniformly colored regions of the 2D image correspond to depth image regions with similar depth values. To evaluate the applicability of superresolution methods which combine a depth and a color image, we implemented a depth-color based superresolution method [42]. For 3D scenes where the depth-color correlation assumption is met, the accuracy of the depth image could indeed be increased by the information from the 2D color image. However, in real-world scenarios we observed that this assumption often is not met

(for example if a 3D object is uniformly colored, or if a planar surface has different colors). In the latter case, the algorithm tends to estimate a relief in the depth image according to the color edges in the 2D image, thus decreasing instead of increasing the accuracy of the depth image.

The second category of superresolution methods (which seek to enhance the depth accuracy by fusing several depth images) is not based on the color-depth correlation assumption and is thus better suited for enhancing the depth measurement accuracy in arbitrary scenarios [43] [44] [45] [46]. While Paolini et al. [47] and Nagesh et al. [48] showed that it is in principle possible to execute superresolution algorithms on a GPU, currently all superresolution algorithms proposed for fusing several depth images are offline algorithms, whose execution can take up to several minutes [49]. These computationally intensive offline algorithms are currently not suited for real-time applications with update rates of up to 30 frames per second. Furthermore, these algorithms pose specific requirements for the acquisition of the input depth images (in view of the feasible rotations and translations of the camera poses at which the input depth images are acquired). Thus, superresolution algorithms fusing several depth images can well be used to reduce the measurement noise of depth images in scenarios where the user can move the camera such that an optimal result is achieved by the superresolution algorithm [46][50]. However, superresolution algorithms are less suited for enhancing the accuracy of depth measurements in a real-time scenario where the user arbitrarily moves a hand-held depth camera around an object to inspect 3D differences from arbitrary viewpoints.

4.2. Enhancing the Depth Measurement Accuracy with 3D Reconstruction based on a Truncated Signed Distance Function

Recently, a depth image based 3D reconstruction algorithm has been proposed which fuses depth images acquired with a hand-held depth camera into a consistent representation of the captured object surface [10][11]. The "KinectFusion" algorithm has met a lot of interest in the computer vision community because it is real-time capable and because it can reconstruct object surfaces with a high precision. The real-time capability is achieved by a highly parallel execution of all calculation steps on the graphics card. Due to this massive parallelization, on a fast graphics card the KinectFusion algorithm can align and fuse depth images acquired by a Kinect at a framerate of up to 30 frames per second. An open source implementation of the KinectFusion algorithm is integrated in the Point Cloud Library [51]. As the KinectFusion algorithm has been described in detail by Newcombe et al. [10] and by Izadi et al. [11], we only briefly sketch its basic principles and refer to the KinectFusion publications [10][11] for details.

To reconstruct the 3D surface of a captured scene, the KinectFusion algorithm discretizes the 3D space into a discrete voxel grid. Each voxel of the grid stores the value of a "Truncated Signed Distance Function" (TSDF) at the center of the voxel. This value of the TSDF at the voxel center represents the distance of the voxel center to the closest reconstructed object surface. Points on the object surface have the distance 0. For 3D points with a non-zero distance, the sign specifies whether the voxel center is inside or outside the object surface. Reconstructing a 3D surface with a TSDF has the advantage that the 3D measurements of several captured depth images are merged for reconstructing the surface, which results in a more accurate surface estimation than the surface estimated from a single depth image. Similar to averaging over time for a static camera position, this merging step reduces measurement noise by fusing the data from several measurements of the same surface position.

Whenever a new depth image is acquired, the new depth image is first aligned with the previously reconstructed 3D model. In the original KinectFusion algorithm, the alignment is based on a point-plane variant of the Iterative Closest Point algorithm [33]. In contrast to the original

KinectFusion implementation, we do not use geometric alignment but rather the camera pose estimated by a high precision measurement arm to align the new depth image with the previous 3D reconstruction. After the alignment of the new depth image with the 3D model has been calculated, the data of the new depth image is merged with the estimation of the 3D surface reconstruction by updating the value of the TSDF for each voxel of the grid. The TSDF is an implicit surface representation which stores only the distances to the closest surface for discrete 3D points in space, but no explicit representation of the surface itself. However, the reconstructed 3D surface can be extracted from the implicit representation, either by the marching cubes algorithm [52] or by ray casting an artificial depth image from a specified (virtual) camera pose.

3D Difference Detection with 3D Surface Reconstruction by the KinectFusion Algorithm. The advantages of the KinectFusion algorithm in comparison to superresolution algorithms are that the KinectFusion algorithm is real-time capable for framerates of up to 30 frames per second and that it does not restrict the way the user can move the depth camera to detect differences between the real object and the 3D model. For these reasons, we integrated the KinectFusion algorithm into the 3D difference detection pipeline (see Figure 1), with the goal to enhance the accuracy of the 3D difference detection by reducing the measurement noise of the captured depth images. To integrate the KinectFusion algorithm with the 3D difference detection pipeline, each depth image acquired by the depth camera is fed into the KinectFusion algorithm. The new depth image is used to update the 3D surface reconstruction. Then, an artificial depth image of the current 3D surface reconstruction is created by ray casting. From the current pose of the depth camera, a ray is casted through each pixel of the virtual image of the depth camera. The depth value of this pixel is calculated by intersecting the view ray with the zero crossing of the TSDF. On a fast graphics card, both the update of the 3D surface reconstruction and the creation of the artificial depth image can be calculated in real-time. Thus, instead of the real depth image captured by the depth camera, the artificial depth image from the KinectFusion algorithm is fed into the 3D difference detection algorithm. Then, the difference detection pipeline is executed the same way as it would for the original depth image acquired by the depth camera. The accuracy enhancement which can be achieved with this integration is quantified in Section 5.2.

5. Quantitative Evaluation

This section provides a quantitative evaluation of the accuracy of depth image based 3D difference detection with different setups (marker based camera pose estimation, pose estimation with a high precision measurement arm and depth measurements from a single depth image vs. fused depth measurements from a reconstructed object surface).

A simulation-based evaluation of the general influence of inaccuracies in the camera pose estimation and the influence of Gaussian depth measurement noise on the overall difference detection accuracy was conducted in a previous publication [17]. While such a simulation provides a first theoretical estimation about how the accuracy of 3D difference detection is influenced by the accuracy of the pose and the 3D measurements, the outcome of the simulation depends on the error model used for the simulation and not all real-world error sources (such as the specific measurement characteristics of real depth cameras) can be modeled. However, to judge the applicability of depth image based 3D difference detection for industrial scenarios, it is necessary to get an estimation of the accuracy which can be achieved in real setups and with real depth cameras. To our knowledge, this section provides the first quantitative evaluation of the measurement

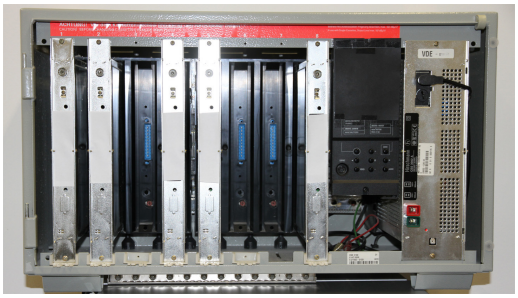
accuracy of depth cameras in hand-held sequences. Previous quantitative evaluations of the measurement accuracy of depth cameras either focused on 3D measurements captured from static camera positions [22] [23] [24] [25] [26] or quantitatively evaluated either a reconstructed 3D model [27] or the accuracy of camera poses estimated based on depth images [53], but did not quantitatively evaluate the measurement accuracy of moving depth cameras.

Depth Cameras. The quantitative evaluation is based on test sequences recorded with state-of-the-art depth cameras (a SwissRanger 4000 time-of-flight depth camera and a Kinect structured-light depth camera). For the SwissRanger 4000 camera, drivers from MesaImaging [21] were used and the data from the Kinect was acquired with the OpenNI interface [37]. Both cameras were calibrated intrinsically. In addition to the intrinsic parameters of the pinhole camera model, the depth and color images were radially undistorted. The depth measurements were used as they were output by the depth cameras, they were not adjusted by a depth calibration. All experiments were conducted with another time-of-flight camera, a CamCube 3.0, as well. This depth camera provides more 3D measurements per depth image than the SwissRanger 4000 (CamCube: 200×200 measurements, SwissRanger: 176×144 measurements). However, the results of the CamCube were much less accurate than the accuracy of both the SwissRanger 4000 and the Kinect depth camera. The measurement accuracy of the CamCube is more influenced by varying factors such as the operational temperature of the camera. These results are in accordance with the experimental evaluation conducted by Piatti [54], who provides a very detailed accuracy analysis of both a SwissRanger 4000 and a CamCube 3.0. In this section, we present the evaluation results of the SwissRanger and the Kinect, which both provided significantly better results than the CamCube 3.0.

Acquisition of Ground Truth Data. The main challenge for conducting a quantitative evaluation of real measurements is the acquisition of ground truth data. For 3D difference detection, a prerequisite for the quantitative evaluation is the availability of a 3D model which exactly corresponds to the real object. For industrial objects, usually no 3D model is available which fulfills this requirement with the necessary precision. For example, the 3D model of the object shown in Figure 4 contains elements which are not part of the real object and vice versa and the 3D model cannot be remodeled easily. The 3D model of the pipes used for 3D difference detection in a previous publication [16] differs even more from the real pipes.

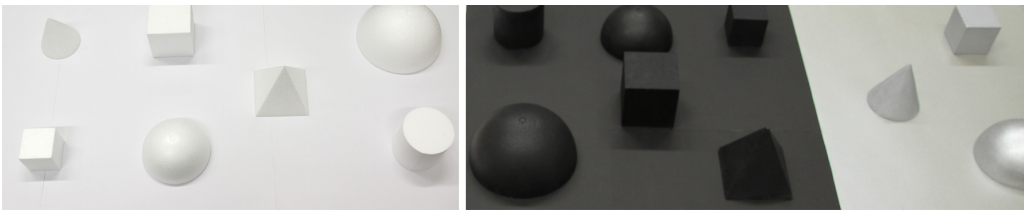
This is why we used the objects shown in Figure 5 for the quantitative evaluation. In comparison to other industrial test objects (such as the one from Figure 4), the 3D model of the industrial, metallic object from Figure 5a matches the real 3D object rather well. Furthermore, we used a 3D object with a curved surface (Figure 5b) for which a very precise 3D model exists and a setup with several convex shapes (hemispheres, cubes, cylinders, cones and pyramids) with different surface colors (white, grey, black). These convex shapes were rigidly attached to planar boards. To create a 3D model of the combined evaluation object, first a separate 3D model was created for each convex shape. Then, the surfaces of the real objects were measured with the measurement tip of a Faro Platinum measurement arm. Finally, a point-triangle mesh variant of the Iterative Closest Point algorithm [33] was used to exactly align each 3D model with the measurement points on the real object. For the alignment, in total 115.000 3D points were measured on the surface of the convex objects with the Faro arm, 80.000 3D points on the object with the curved surface and 24.000 3D points on the outer surface of the metallic object. After the alignment, the average distance between the surface of the 3D model and the measurements on

the surface of the real object was 0.3mm for the convex objects, 0.2mm for the object with the curved surface and 0.1mm for the metallic object.



(a) Industrial, metallic evaluation object.

(b) Evaluation object with curved surface.



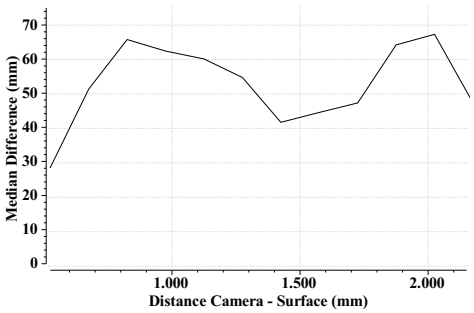
(c) Convex evaluation objects with different surface properties.

Figure 5: Evaluation objects for quantitative ground truth evaluation.

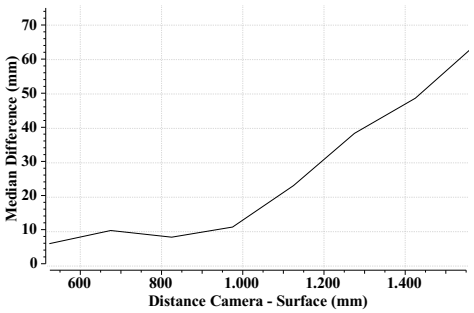
For each evaluation object, a depth image sequence was captured both with a SwissRanger 4000 depth camera and with a Kinect depth camera. To acquire the evaluation sequences with a constant framerate, the framerate was limited to 10 fps (the SwissRanger depth camera automatically adjusts its integration time to the captured scene, so a higher framerate can only be achieved if the integration time is less than 100ms). Each Kinect depth image sequence consists of 500 to 800 frames, each SwissRanger 4000 depth image sequence of 4000 to 7000 frames. More images were captured with the SwissRanger depth camera to get an equivalent number of 3D measurements - while the Kinect outputs 640×480 depth values per frame, the SwissRanger has a resolution of 176×144 depth values. To compare the depth measurements with the ground truth data (the distance from the camera to the surface of the 3D model), the 3D model was rendered from the current pose of the depth camera. Then, the values of the rendering depth buffer were compared to the depth values measured by the depth camera. For each captured sequence, 10 to 38 million 3D measurements on the surface of the evaluation object were compared to their corresponding ground truth values from the rendered depth buffer.

5.1. Experiment 1: Enhancing the Accuracy of 3D Difference Detection with a High Precision Pose Estimation

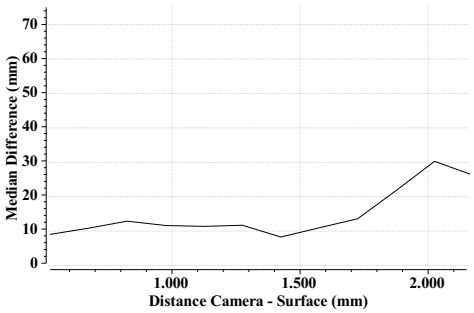
The purpose of this experiment was to quantify the 3D difference detection accuracy which can be achieved when the camera pose is estimated either image based (with a marker tracker) or with a mechanical measurement arm. For this experiment, the evaluation object with the curved surface from Figure 5b was used. In the marker tracking mode, the pose was estimated from the three square markers attached above the evaluation object. The Kinect contains both a built-in



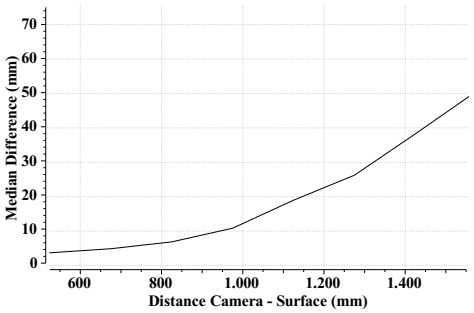
(a) SwissRanger 4000, marker based pose estimation.



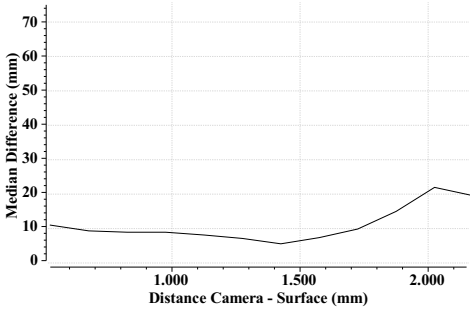
(b) Kinect, marker based pose estimation.



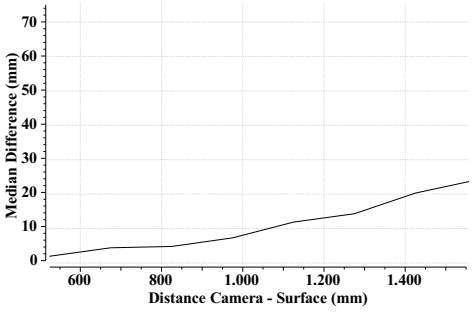
(c) SwissRanger 4000, Faro pose estimation.



(d) Kinect, Faro pose estimation.



(e) SwissRanger 4000, Faro pose estimation, diff. detection based on reconstructed 3D model.



(f) Kinect, Faro pose estimation, difference detection based on reconstructed 3D model.

Figure 6: 3D difference detection accuracy for marker based pose estimation, measurement arm pose estimation and measurement arm based pose estimation in combination with the 3D surface reconstruction algorithm. Evaluation object: curved surface (Figure 5b).

color and a depth camera. For the marker-based pose estimation with the Kinect, the relative transformation between the color and the depth camera was calculated in an offline calibration procedure. For evaluating the accuracy of 3D difference detection with a marker tracker for the Kinect, the markers were tracked with the built-in 640×480 color camera and the pose of the depth camera was inferred from the pose of the color camera by adding the previously calculated relative transformation to the pose of the color camera. In contrast to the Kinect, the SwissRanger depth camera does not contain an additional color camera, but measures an intensity (grey) value for each captured depth measurement. The markers were tracked with this 176×144 intensity image. Figure 6a and Figure 6b show the overall accuracy of the 3D difference detection when the pose of the depth camera is estimated with a marker tracker.

To estimate the accuracy of 3D difference detection when the pose of the depth camera is measured with a Faro measurement arm, in a second step the same sequences were evaluated with a measurement arm based pose estimation. The pose of the depth camera was inferred from the pose of the measurement arm by adding the previously calculated hand-eye calibration (between the tip of the measurement arm and the depth camera) to the measured pose of the Faro arm. Figure 6c and Figure 6d visualize the overall accuracy of the 3D difference detection when the pose is estimated with a Faro measurement arm. The measurement arm based pose estimation increases the accuracy for both depth cameras. This effect is more pronounced with the SwissRanger depth camera because the marker based pose estimation of the SwissRanger is less accurate than the marker based pose estimation of the Kinect camera. This is due to the lower resolution of the SwissRanger's 2D image used for the marker based camera pose estimation. The accuracy of the 3D difference detection with the Kinect depth camera is increased as well.

5.2. Experiment 2: Enhancing the Accuracy of 3D Difference Detection with a 3D Surface Reconstruction Algorithm

This experiment was conducted to evaluate whether the accuracy of 3D difference detection can further be improved if the 3D difference detection is based a reconstructed 3D model of the captured object (instead of using a single depth image for the 3D difference detection). For this purpose, the TSDF reconstruction step of the KinectFusion algorithm [10][11] was integrated in the 3D difference detection pipeline as described in Section 4.2.

For each new depth image acquired by the depth camera, the new depth image is fed into this algorithm to reconstruct the 3D surface of the captured object. In order to avoid the problems described in section 3.2, the external pose acquired from the measurement arm is used to align the new depth measurements with the previously reconstructed 3D model. This replaces the geometric alignment step of the original KinectFusion algorithm. After the reconstruction algorithm has updated its 3D surface estimation with the data of the new depth image, an artificial depth image is generated from the reconstructed 3D surface by raycasting. Instead of the depth image currently captured by the depth camera, this raycasted depth image was fed back into the 3D difference detection pipeline and compared to the provided reference 3D model of the real object. Figure 6e and Figure 6f show the overall accuracy of the 3D difference detection which is achieved when the camera pose is estimated with a Faro measurement arm in combination with the 3D surface reconstruction algorithm.

Figure 7 shows the accuracy of the 3D difference detection for the setup from Figure 4, both with and without the integration of the 3D surface reconstruction algorithm. The color scale on the right visualizes the color encoding of the measured differences (red: this pixel was measured closer than represented by the 3D model, yellow: this pixel was measured to be farther away than modeled). For example, the real gearshift differs from the modeled gear (both in view of its

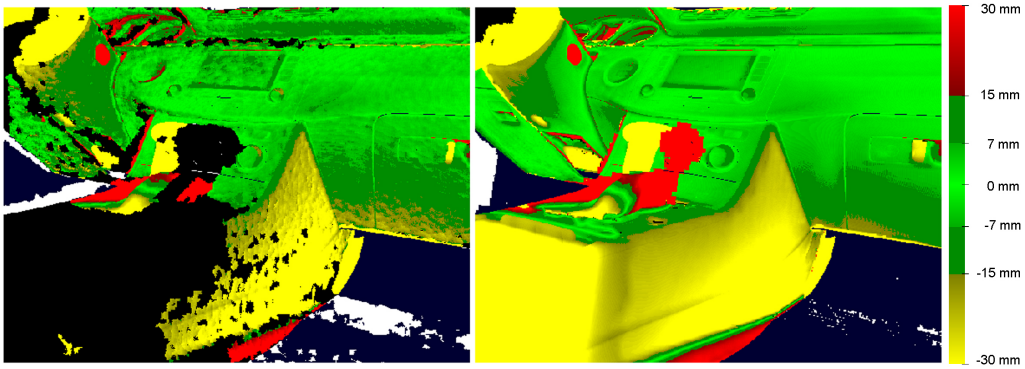


Figure 7: 3D difference detection based on single depth image (left) and with reconstructed 3D model (right).

position and its shape) and the wheel is part of the 3D model, but not of the built mockup. A pixel is colored in black if the depth camera could not capture a depth measurement at this position (and thus, no 3D difference could be calculated). The pixels colored in dark blue are not part of the 3D model. The 3D surface reconstruction algorithm not only reduces the measurement inaccuracies on large parts of the measured surface but also reduces the areas for which no 3D difference could be calculated at all.

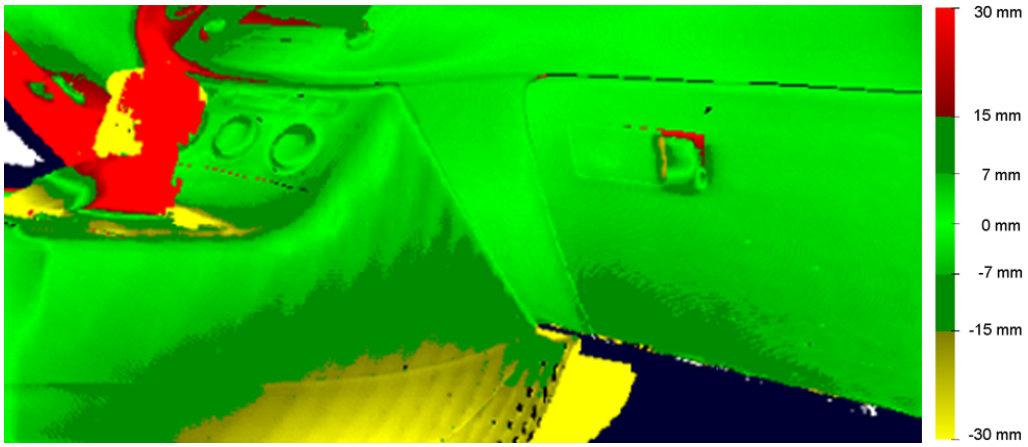


Figure 8: The reconstruction of the side of the center part improves step-by-step when captured from a more orthogonal point of view.

In Figure 7, the right side of the central element is colored in yellow although it does not differ much from the provided 3D model. This is because the previously recorded depth images were captured from a frontal view of the whole object. In the frame visualized in Figure 7, the camera has only just begun to move to the right. When the camera continues to be moved to the right (and to capture the right side of the central element from a more orthogonal point of view), the reconstruction of this part of the captured object improves frame-by-frame (see Figure 8).

Enhancing the Accuracy with 3D Reconstruction, but without High Precision Pose Estimation. Measurement arms provide a high precision pose estimation, but are significantly more expensive than a depth camera. Therefore, we also evaluated the accuracy of a setup which does not require a measurement arm. In this setup, the camera pose is estimated with a marker tracker and the accuracy of the 3D difference detection is enhanced with the 3D surface reconstruction algorithm.

Table 1 provides the numerical values of the 3D difference detection accuracy with the Kinect depth camera for the marker based pose estimation and for the pose estimation with a Faro measurement arm, both with and without the accuracy enhancement by the 3D surface reconstruction. Both the measurement arm based pose estimation and the 3D reconstruction algorithm improve the overall accuracy of the 3D difference detection. The best accuracy is achieved when these two approaches are combined. For the Kinect, the setup "marker pose with 3D reconstruction" also provides good results with a much more cost-effective hardware solution. The main drawback of the setup "marker pose with 3D reconstruction" is not so much the accuracy, but rather the fact that the marker always needs to be visible in the depth camera image.

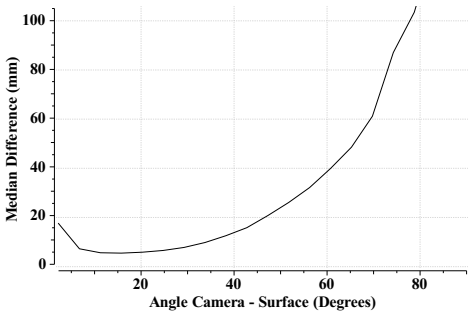
Distance Kinect camera - surface	**Marker pose without 3D rec.**	**Marker pose with 3D rec.**	**Faro pose without 3D rec.**	**Faro pose with 3D rec.**
450-599	6.54	7.76	3.70	1.96
600-749	10.34	10.71	4.88	4.41
750-899	8.40	6.50	6.87	4.80
900-1049	11.34	8.54	10.84	7.30
1050-1199	23.39	13.37	18.97	11.88
1200-1349	38.56	22.81	26.24	14.31
1350-1499	48.78	39.85	38.26	20.31
1500-1649	64.49	48.35	50.58	24.11

Table 1: Median difference between 3D measurements and ground truth (numerical values of Figure 6 for difference detection with a Kinect depth camera). Additionally, the results of the setup "marker pose with 3D reconstruction" are provided, which can be used without a measurement arm. All values are in mm.

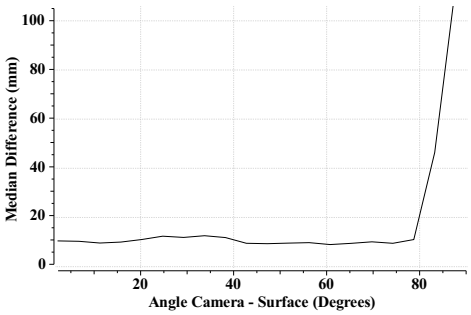
In contrast to the Kinect, the 3D difference detection of the SwissRanger 4000 could not be improved by the 3D reconstruction algorithm when a marker tracker was used for the pose estimation. As the pose can only be estimated very coarsely with a marker tracker on the low resolution intensity image of such a depth camera, most of the errors visualized in Figure 6a were caused by errors in the pose estimation. Thus, they can not be smoothed out by the 3D surface reconstruction algorithm, which is based on the pose estimation to reconstruct the 3D surface.

5.3. Experiment 3: Influence of the Angle on the Measurement Accuracy

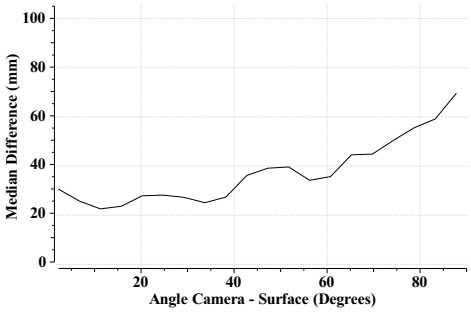
To quantify the effect of the measurement angle on the measurement accuracy, the accuracy of the 3D difference detection was evaluated as a function of the measurement angle. First, the surface normal of the object surface was calculated for each pixel of the depth buffer image acquired from the rendered 3D model. Then, the angle between the surface normal and the view ray from the optical center of the depth camera through the current pixel was calculated. This angle varies between 0° and 90°. Figure 9 visualizes the accuracy of the difference detection as a function of the measurement angle. The influence of the measurement angle is plotted for three different setups (the evaluation objects from Figure 5a, Figure 5b and the white convex objects from Figure 5c). For both cameras, the measurement accuracy decreases for large angles of more than about 60°.



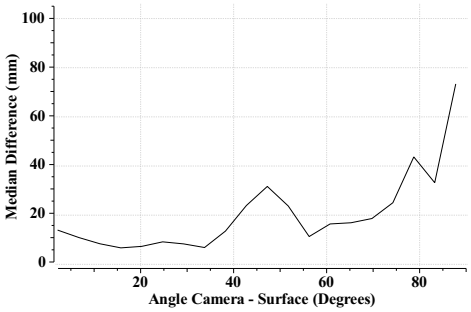
(a) SwissRanger 4000, Faro pose estimation, curved surface.



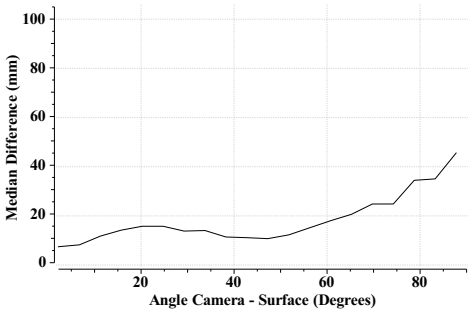
(b) Kinect, Faro pose estimation, curved surface.



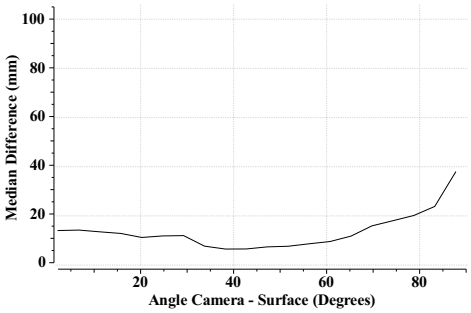
(c) SwissRanger 4000, Faro pose estimation, metallic object.



(d) Kinect, Faro pose estimation, metallic object.

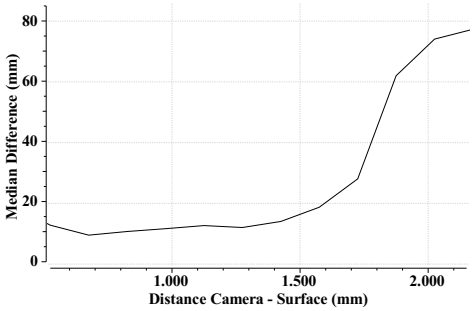


(e) SwissRanger 4000, Faro pose estimation, white convex objects.

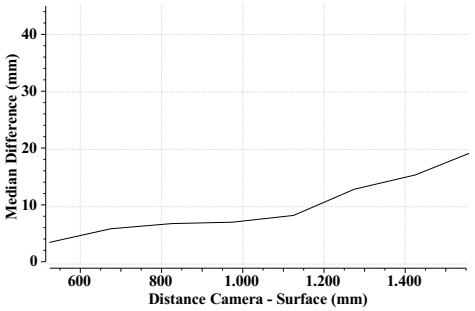


(f) Kinect, Faro pose estimation, white convex objects.

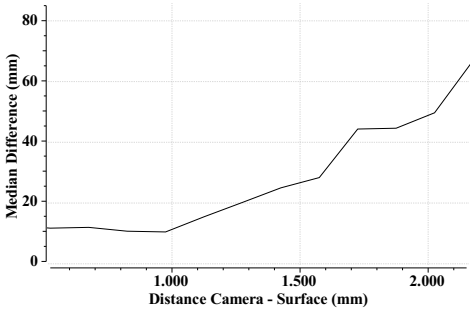
Figure 9: Dependency of the measurement accuracy on the measurement angle between the camera and the object surface.



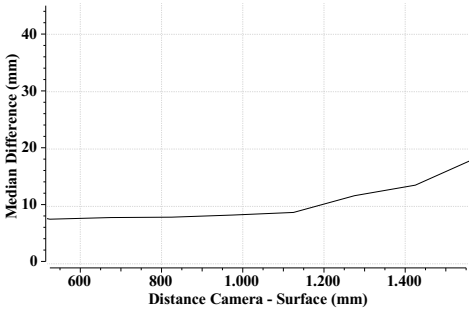
(a) SwissRanger 4000, Faro pose estimation, white convex objects.



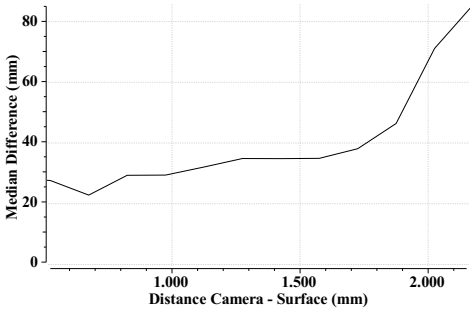
(b) Kinect, Faro pose estimation, white convex objects.



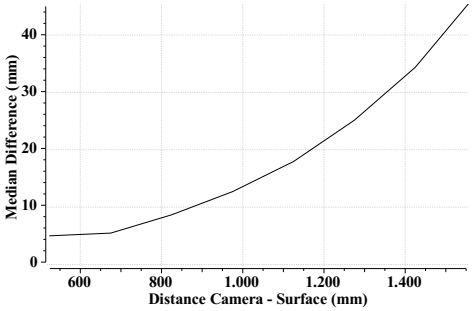
(c) SwissRanger 4000, Faro pose estimation, black convex objects.



(d) Kinect, Faro pose estimation, black convex objects.



(e) SwissRanger 4000, Faro pose estimation, metallic object.



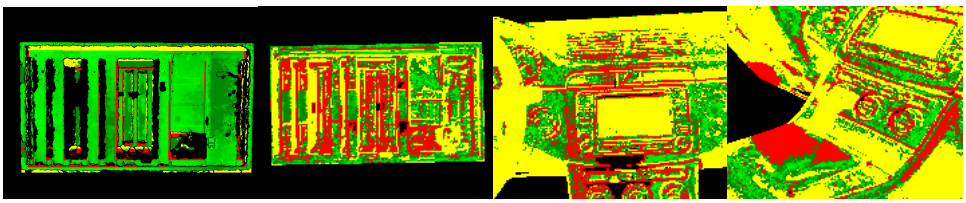
(f) Kinect, Faro pose estimation, metallic object.

Figure 10: Dependency of the measurement accuracy on different object surfaces (white, black and specular metallic).

5.4. Experiment 4: Influence of Object Surface Properties on the Measurement Accuracy

In industrial applications, the surfaces of objects often are metallic or have dark colors. Such surfaces are more difficult to measure with depth cameras than diffuse, light surfaces. To quantify the effect of different surface properties on the accuracy of the 3D difference detection, we evaluated the 3D difference detection accuracy for three objects with different surface properties. The accuracy was evaluated for convex shapes with a white respectively black surface (Figure 5c) and for a metallic, industrial object (Figure 5a). Figure 10 visualizes the results of this evaluation. For the SwissRanger depth camera, up to a distance of one meter, the accuracy of the difference detection is similar for the white and black shapes. However, for distances larger than one meter, the accuracy decreases faster for the black surfaces. Even for close distances of the camera to the object, the measurement accuracy of the SwissRanger depth camera is low for the metallic object. This effect can be explained by the measurement principle of the time-of-flight depth camera: The light emitted by the time-of-flight camera is reflected multiple times by the metallic surface before it gets reflected to the depth camera. This increases the time it takes until the light emitted by the camera is captured by the camera sensor. Due to this prolonged time-of-flight of the emitted light, the depth camera overestimates the distance to the captured object.

In contrast to the time-of-flight depth camera, the accuracy of the structured light Kinect depth camera gets less affected by the metallic surface of the industrial object. Although the surface of this object is strongly specular, for close distances of the Kinect depth camera to the surface, the distances to the object are measured with a high precision. Just as for the other evaluation objects, the measurement accuracy of the Kinect depends much more on the distance of the camera to the objects than on the surface properties of the captured objects. Figure 11 shows the difference visualization of metallic and black surfaces captured with a Kinect and a SwissRanger depth camera (see also Figure 7 for comparison with the depth image acquired by the Kinect). For such surfaces, the measurements acquired by the SwissRanger depth camera are much noisier than those acquired by the Kinect.



(a) Kinect (b) SwissRanger 4000 (c) SwissRanger 4000 (d) SwissRanger 4000

Figure 11: Objects with metallic and black surfaces captured with a SwissRanger 4000 and a Kinect depth camera (same objects as in Figure 5a and 7). The color encoding of the differences is the same as in Figure 7.

6. Discussion of the Evaluation Results and Future Work

In this paper, we have presented a 3D difference detection pipeline which enhances the accuracy of depth image based 3D difference detection with two approaches: First, the precision of the depth camera's pose estimation is improved by coupling the depth camera with a high precision industrial measurement arm. Second, the influence of the depth measurement noise

is reduced by integrating a 3D surface reconstruction algorithm in the 3D difference detection pipeline. The effect of both enhancements on the achievable accuracy of the 3D difference detection was quantified in the ground-truth based quantitative evaluation, both for a time-of-flight depth camera (SwissRanger 4000) and a structured light depth camera (a Kinect depth camera). The evaluation was conducted with several evaluation objects and different surface properties.

The accuracy of the 3D difference detection which can be achieved with the SwissRanger 4000 time-of-flight depth camera depends strongly on the surface of the captured object. For example, the accuracy strongly decreases if the captured object has a metallic surface. In contrast, the measurements acquired by the Kinect depth camera are more independent of the surface properties of the captured scene and even the metallic surfaces can be captured well with a Kinect. Thus, the Kinect seems to be better suited for industrial scenarios with metallic objects. At close distances of the Kinect depth camera to the captured surface, the combination of the measurement arm based pose estimation and the 3D reconstruction algorithm reduces the overall 3D difference detection error to a few millimeters (about 1.73mm at a distance of 45-60cm and 4.68mm at a distance of 75-90cm). This is an important step towards the applicability of depth camera based 3D difference detection in industrial scenarios.

A drawback of the current approach is that the 3D difference detection accuracy decreases very strongly when the distance between the Kinect and the captured surface increases. From a visualization of the measured distances and the ground truth distances, this effect seems to be caused by systematic measurement errors. When the distance between the Kinect and the measured surface increases, the Kinect seems to consistently overestimate the distance to the measured surface for all captured depth values. For this reason, as a next step, we will conduct a depth calibration of the Kinect depth camera (which corrects the measured depth values as a function of the measured distance and should thus reduce the systematic measurement errors). Then, we will evaluate how much the 3D difference detection can be further improved by such a depth calibration.

Another important aspect for future work is a closer integration of the 3D difference detection pipeline and the 3D surface reconstruction algorithm (KinectFusion). Currently, only the depth image is exchanged between the 3D surface reconstruction and the other components of the 3D difference detection pipeline. The KinectFusion algorithm estimates the offset between the camera pose of the first depth image and the current depth image by geometrically aligning the depth images with the reconstructed 3D surface. This approach has the drawbacks described in Section 3.2: The alignment diverges if the camera is moved close to the captured surface, if it is moved too far away from the captured surface or if the surface of the captured object has only few distinct geometric features. Therefore, we will evaluate whether the robustness and accuracy can be further improved if the pose acquired by the mechanical measurement arm is used in the KinectFusion algorithm instead of the geometric alignment step.

References

[1] T. Oggier, F. Lustenberger, N. Blanc, Miniature 3d tof camera for real-time imaging, in: Perception and Interactive Technologies 2006, pp. 212–216.

[2] A. Kolb, E. Barth, R. Koch, R. Larsen, Time-of-flight sensors in computer graphics, in: Proc. Eurographics (State-of-the-Art Report) 2009, pp. 119–134.

[3] Z. Zalevsky, A. Shpunt, A. Maizels, J. Garcia, Method and system for object reconstruction, 2007.

[4] G. Sansoni, M. Trebeschi, F. Docchio, State-of-the-art and applications of 3d imaging sensors in industry, cultural heritage, medicine, and criminal investigation, Sensors 9 (2009) 568–601.

[5] Z. Bi, L. Wang, Advances in 3d data acquisition and processing for industrial applications, Robotics and Computer-Integrated Manufacturing 26 (2010) 403 – 413.

[6] E. Prados, O. Faugeras, Shape from shading, in: N. Paragios, Y. Chen, O. Faugeras (Eds.), Handbook of Mathematical Models in Computer Vision, Springer US, 2006, pp. 375–388.
[7] Faro Focus3D, 2013. Http://www.faro.com/focus. Date of access: 02/2013.
[8] Leica laser scanner, 2013. Http://hds.leica-geosystems.com. Date of access: 02/2013.
[9] Faro ScanArm, 2013. Http://measuring-arms.faro.com/scanarm. Date of access: 02/2013.
[10] R. A. Newcombe, S. Izadi, O. Hilliges, D. Molyneaux, D. Kim, A. J. Davison, P. Kohli, J. Shotton, S. Hodges, A. Fitzgibbon, Kinectfusion: Real-time dense surface mapping and tracking, in: Proceedings of the 2011 10th IEEE International Symposium on Mixed and Augmented Reality, ISMAR 2011, pp. 127–136.
[11] S. Izadi, D. Kim, O. Hilliges, D. Molyneaux, R. Newcombe, P. Kohli, J. Shotton, S. Hodges, D. Freeman, A. Davison, A. Fitzgibbon, Kinectfusion: real-time 3d reconstruction and interaction using a moving depth camera, UIST 2011, pp. 559–568.
[12] P. Georgel, P. Schroeder, S. Benhimane, S. Hinterstoisser, M. Appel, N. Navab, An industrial augmented reality solution for discrepancy check, in: ISMAR 2007: Proceedings of the 6th IEEE and ACM International Symposium on Mixed and Augmented Reality, pp. 1–4.
[13] P. Georgel, P. Schroeder, N. Navab, Navigation tools for viewing augmented cad models, IEEE Computer Graphics and Applications 29 (2009) 65–73.
[14] P. Georgel, S. Benhimane, J. Sotke, N. Navab, Photo-based industrial augmented reality application using a single keyframe registration procedure, in: ISMAR 2009: Proceedings of the 8th IEEE and ACM International Symposium on Mixed and Augmented Reality, pp. 187–188.
[15] S. Webel, M. Becker, D. Stricker, H. Wuest, Identifying differences between cad and physical mock-ups using ar, in: ISMAR 2007: Proceedings of the Sixth IEEE and ACM International Symposium on Mixed and Augmented Reality, pp. 281–282.
[16] S. Kahn, H. Wuest, D. Stricker, D. W. Fellner, 3d discrepancy check via augmented reality, in: 9th IEEE International Symposium on Mixed and Augmented Reality (ISMAR) 2010, pp. 241–242.
[17] S. Kahn, Reducing the gap between augmented reality and 3d modeling with real-time depth imaging, Virtual Reality (2011) 1–13. Http://dx.doi.org/10.1007/s10055-011-0203-0.
[18] F. Bosché, J. Teizer, C. T. Haas, C. H. Caldas, Integrating data from 3d cad and 3d cameras for real-time modeling, in: Proceedings of Joint International Conference on Computing and Decision Making in Civil and Building Engineering 2006, pp. 37–46.
[19] F. Bosché, Automated recognition of 3d cad model objects in dense laser range point clouds, Ph.D. thesis, University of Waterloo, 2008.
[20] F. Bosché, Automated recognition of 3d cad model objects in laser scans and calculation of as-built dimensions for dimensional compliance control in construction, J. Adv. Eng. Inf. 24 (2010) 107–118.
[21] Mesa Imaging, 2013. Http://www.mesa-imaging.ch. Date of access: 02/2013.
[22] S. Fuchs, G. Hirzinger, Extrinsic and depth calibration of tof-cameras, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2008, pp. 1–6.
[23] H. Rapp, M. Frank, F. Hamprecht, B. Jahne, A theoretical and experimental investigation of the systematic errors and statistical uncertainties of time-of-flight-cameras, Int. J. Intell. Syst. Technol. Appl. 5 (2008) 402–413.
[24] M. Lindner, I. Schiller, A. Kolb, R. Koch, Time-of-flight sensor calibration for accurate range sensing, Computer Vision and Image Understanding 114 (2010) 1318–1328.
[25] A. Belhedi, S. Bourgeois, V. Gay-Bellile, P. Sayd, A. Bartoli, K. Hamrouni, Non-parametric depth calibration of a tof camera, in: ICIP 2012, pp. 549–552.
[26] K. Khoshelham, S. O. Elberink, Accuracy and resolution of kinect depth data for indoor mapping applications, Sensors 12 (2012) 1437–1454.
[27] J. C. K. Chow, K. D. Ang, D. D. Lichti, W. F. Teskey, Performance analysis of a low-cost triangulation-based 3d camera: Microsoft kinect system, ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XXXIX-B5 (2012) 175–180.
[28] PrimeSense, 2013. Http://www.primesense.com. Date of access: 02/2013.
[29] H. Gonzalez-Jorge, B. Riveiro, E. Vazquez-Fernandez, J. Martinez-Sanchez, P. Arias, Metrological evaluation of microsoft kinect and asus xtion sensors, J. Measurement (2013) http://dx.doi.org/10.1016/j.measurement.2013.01.011.
[30] G. Bleser, M. Becker, D. Stricker, Real-time vision-based tracking and reconstruction, J. Real Time Image Proc. 2 (2007) 161–175.
[31] J. Shi, C. Tomasi, Good features to track, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR'94), pp. 593–600.
[32] H. Wuest, Efficient line and patch feature characterization and management for real-time camera tracking, Ph.D. thesis, TU Darmstadt, 2008.
[33] P. Besl, N. McKay, A method for registration of 3-d shapes, in: IEEE Transactions on Pattern Analysis and Machine Intelligence 1992, volume 14(2), pp. 239–256.

[34] S. Kahn, A. Kuijper, Fusing real-time depth imaging with high precision pose estimation by a measurement arm, in: 2012 International Conference on Cyberworlds (CW), pp. 256–260.
[35] R. Y. Tsai, R. K. Lenz, A new technique for fully autonomous and efficient 3d robotics hand-eye calibration, in: Proceedings of the 4th international symposium on Robotics Research 1998, pp. 287–297.
[36] K. H. Strobl, G. Hirzinger, Optimal hand-eye calibration, in: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) 2006, pp. 4647–4653.
[37] OpenNI, Openni framework, 2013. Http://www.openni.org/. Date of access: 02/2013.
[38] S. Farsiu, M. Robinson, M. Elad, P. Milanfar, Fast and robust multiframe super resolution, IEEE Transactions on Image Processing 13 (2004) 1327 –1344.
[39] V. Garro, C. d. Mutto, P. Zanuttigh, G. M. Cortelazzo, A novel interpolation scheme for range data with side information, in: Proceedings of the 2009 Conference for Visual Media Production, CVMP '09, pp. 52–60.
[40] B. Langmann, K. Hartmann, O. Loffeld, Comparison of depth super-resolution methods for 2d/3d images, Int. J. Comput. Inf. Syst. Ind. Manag. Appl. 3 (2011) 635 –645.
[41] J. Lu, D. Min, R. Pahwa, M. Do, A revisit to mrf-based depth map super-resolution and enhancement, in: 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 985 –988.
[42] J. Diebel, S. Thrun, An application of markov random fields to range sensing, in: NIPS'05, pp. 291–298.
[43] T. Edeler, K. Ohliger, S. Hussmann, A. Mertins, Super resolution of time-of-flight depth images under consideration of spatially varying noise variance, in: 2009 16th IEEE International Conference on Image Processing (ICIP), pp. 1185 –1188.
[44] Y. J. Kil, N. Amenta, B. Mederos, Laser scanner super-resolution, in: Eurographics Symposium on Point-Based Graphics 2006, pp. 9–15.
[45] Y. M. Kim, K. Zhu, Super-resolution 3d multiview reconstruction using time-of-flight depth sensors, Technical Report, Stanford University, 2008.
[46] S. Schuon, C. Theobalt, J. Davis, S. Thrun, Lidarboost: Depth superresolution for tof 3d shape scanning, in: CVPR, IEEE, 2009, pp. 343–350.
[47] A. L. Paolini, F. Ortiz, D. K. Price, K. E. Spagnoli, Development of a gpu-accelerated super resolution solver, in: Proc. SPIE 7348 (2009), pp. 1–11.
[48] P. Nagesh, R. Gowda, B. Li, Fast gpu implementation of large scale dictionary and sparse representation based vision problems, in: Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on, pp. 1570 –1573.
[49] Y. Cui, D. Stricker, 3d shape scanning with a kinect, in: ACM SIGGRAPH 2011 Posters, pp. 57–57.
[50] Y. Cui, S. Schuon, C. Derek, S. Thrun, C. Theobalt, 3d shape scanning with a time-of-flight camera, in: Proc. of IEEE CVPR 2010, pp. 1173–1180.
[51] R. B. Rusu, S. Cousins, 3d is here: Point cloud library (pcl), in: International Conference on Robotics and Automation 2011, pp. 1–4.
[52] W. E. Lorensen, H. E. Cline, Marching cubes: A high resolution 3D surface construction algorithm, in: Computer Graphics (Proceedings of SIGGRAPH 1987), volume 21, pp. 163–169.
[53] S. Lieberknecht, A. Huber, S. Ilic, S. Benhimane, Rgb-d camera-based parallel tracking and meshing, in: Proceedings of the 2011 10th IEEE International Symposium on Mixed and Augmented Reality, ISMAR 2011, pp. 147–155.
[54] D. Piatti, Time-of-Flight cameras: tests, calibration and multi-frame registration for automatic 3D object reconstruction, Ph.D. thesis, Politecnico di Torino, Italy, 2011.