Multi-view ToF Fusion for Object Detection in Industrial Applications

Inge Coudron and Toon Goedemé
EAVISE Research Group, KU Leuven, Jan Pieter De Nayerlaan 5, Sint-Katelijne-Waver, Belgium
{inge.coudron, toon.goedeme}@kuleuven.be

Keywords: Extrinsic Calibration, Multi-sensor, Object Detection.

Abstract: The use of time-of-flight (ToF) cameras in industrial applications has become increasingly popular due to the camera's reduced cost and its ability to provide real-time depth information. Still, one of the main drawbacks of these cameras has been their limited field of view. We therefore propose a technique to fuse the views of multiple ToF cameras. By mounting two cameras side by side and pointing them away from each other, the horizontal field of view can be artificially extended. The combined views can then be used for object detection. The main advantages of our technique is that the calibration is fully automatic and only one shot of the calibration target is needed. Furthermore, no overlap between the views is required.

1 INTRODUCTION

Object detection remains an important challenge in industry. In many of these applications, a large scene area needs to be covered. Hence a camera with a wide field of view is usually required. Since the field of view of a single ToF camera is limited, multiple cameras must be combined. This requires to first calibrate the relative poses (i.e. extrinsic parameters) of the cameras.

Once the data from the different cameras is transformed into a common reference frame, it can be fed to the object detection framework. A popular approach to the 3D object detection problem is to exploit range images (Bielicki and Sitnik, 2013). These images make data processing significantly faster, as they convert the most time-consuming tasks (e.g., nearest neighbor search) from a 3D space into a 2D space. We will therefore render the registered point clouds with a virtual camera to simulate a depth sensor.

In this paper, we present a convenient external calibration method for a multi-ToF system. That is to say, the human interaction and export knowledge required for the calibration is kept to a minimum. The views from the different ToF cameras can be merged into an extended range image usable for 3D object detection. The remainder of this paper is organized as follows. Firstly, the Related Work section provides an overview of existing calibration techniques. Section 3 introduces our approach for multi-view TOF fusion. Experiments in section 4 show the accuracy in calibration. Finally, a short conclusion is given.

2 RELATED WORK

The calibration of multiple cameras is a well-studied problem in computer vision. The most common method for calibrating conventional intensity cameras is to use a checkerboard which is observed at different positions and orientations within the cameras shared field of view (Zhang, 2000). Given the image coordinates of the reference points (i.e., the checkerboard corners) and the geometry of the checkerboard (i.e., the number of squares and the square dimension), the camera parameters can be estimated using a closed form solution w.r.t. the pinhole camera model. An iterative bundle adjustment algorithm can then be used to refine the parameters. The same standard technique could be used for ToF cameras as well, as they provide an amplitude image associated with each range image. However, the low resolution of the amplitude images makes it difficult to detect the checkerboard corners reliably resulting in inaccurate calibration.

To overcome this limitation, other methods have been proposed that work directly on 3D shapes. Auvinet et al. (Auvinet et al., 2012), for example, use the intersection points of triplets of planes as reference points. The equation of each plane can be calculated by using a singular value decomposition of points lying on the plane. Given the sets of corresponding reference points, the rigid body transformation between the pair of cameras is estimated in a least square sense. Another method presented by Ruan et al. (Ruan and Huber, 2014), uses the centers of a spherical calibration target as reference points. The

Coudron I. and GoedemÃl' T.
Multi-view ToF Fusion for Object Detection in Industrial Applications.
DOI: 10.5220/0006136802030207
In *Proceedings of the 12th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2017)*, pages 203-207
ISBN: 978-989-758-225-7

spherical target has the advantage that it is rotation invariant. However, the extraction of the center of the sphere from a noisy point cloud is less robust, since the active illumination will be mostly scattered away by the spherical surface. Furthermore, note that these methods still require a significant amount of human operation, since the calibration target must be moved and matched in many different positions.

Alternatively, a method that takes only one pair of 3D images is the Iterative Closest Point (ICP) algorithm (Besl and McKay, 1992). The ICP algorithm has been widely used for 3D registration. The rigid body transformation between two point clouds is estimated by minimizing the distance from a point in one cloud to the closest point in the other cloud. A popular variant of this method minimizes the distance between a point and the tangent plane at its correspondence point instead. The point-to-plane error metric usually performs better in structured environments (Low, 2004). One of the advantages of this algorithm is that in contrast to the first algorithm, it does not rely on local feature extraction. Unfortunately, the ICP algorithm requires sufficient overlap between the point clouds to succeed (Chetverikov et al., 2002). Hence, at first sight, it might not seem like an appropriate method for the artificial extension of a ToF camera's field of view addressed in this work. We will however show that using a proper calibration target, the method contributes to an easy to use calibration tool.

3 MATERIALS AND METHODS

3.1 Camera Set-up

The camera set-up is depicted in 1. Two IFM Efector O3D303 ToF cameras are mounted side by side and pointing away from each other to artificially extend the field of view as can be seen in figure 2. The red triangle depicts the field of view that would have been covered if a single ToF camera was placed in the middle. As can be seen in this figure, a small area at the bottom is not covered. This is not necessarily a problem since the minimum operating distance must be respected anyway. With two cameras, the grey area that is covered depends on the angle between the cameras. Increasing the angle, further extends the area covered. However, this also implies that the uncovered area in the middle enlarges. Hence a tradeoff must be made depending on the desired operating range.

To be able to fuse the data from the different cameras, it is important that the images are captured synchronously. This can be achieved by cascading the cameras via hardware trigger. The first camera will automatically trigger the second camera after completion of the image capture. However, if both cameras are operating on the same active illumination frequency measurement errors may occur due to mutual interference from simultaneous exposure (see Fig. 3). By setting the cameras on a different frequency channel the occurrence of measurement errors can be reduced. Both cameras are connected to a single GigE port through a switch.

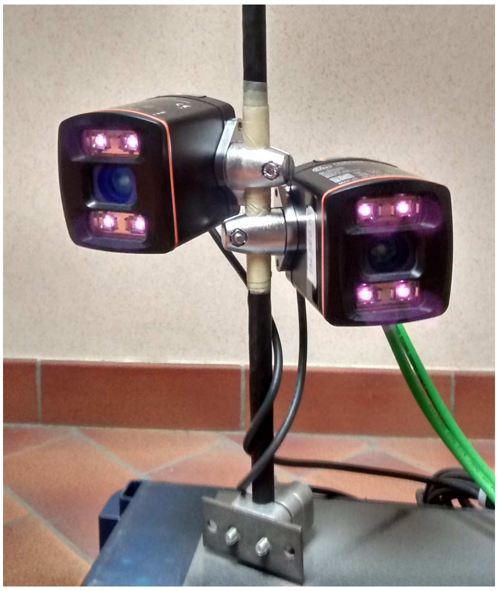

Figure 1: Camera set-up.

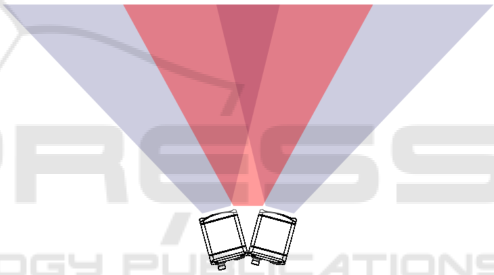

Figure 2: Artificially extended field of view.

3.2 External Calibration

To determine the rigid body transformation between two point clouds, 6 degrees of freedom (DoF) need to be eliminated. In theory, a set of 4 non-coplanar reference points is sufficient. Nevertheless, it is best to use as many points as possible to increase the reliability of the transformation found. As such, a calibration target is defined using geometric primitives

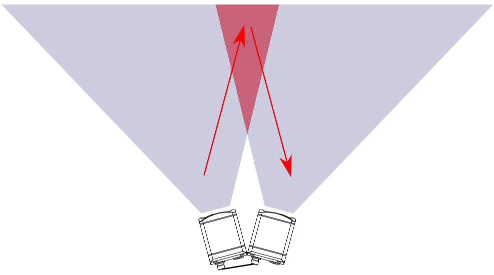

Figure 3: Mutual interference by ToF cameras during simultaneous operation with the same active illumination frequency.

that eliminates all DoFs. Possible primitives are a sphere eliminating 3 DoF (viz. three translations), a plane eliminating 3 DoF as well (viz. one translation and two rotations) and others. In this work, we will use a calibration target with multiple planar regions. The motivation for this selection is that planes can be acquired more reliably with a ToF camera than shapes with varying normals such as spheres.

The main idea behind our approach is that we do not directly register the two ToF views with each other. Instead, each camera individually registers only to the observed part of the calibration target using the ICP algorithm. Since both parts of the calibration target are defined in the same coordinate system, the point clouds are transformed into a common reference frame and the extrinsic parameters between the cameras can be derived. The proposed calibration target can be seen in Fig. 4. The calibration target is split into two parts, one that is observed by the left camera and one by the right camera (see Fig. 5).

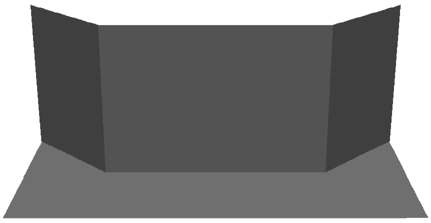

Figure 4: CAD model of the calibration target.

A few considerations still need to be addressed. First of all, the ICP algorithm is a local registration method, meaning that an initial estimation of the global transformation is necessary to obtain a good result. Chances are otherwise that the algorithm will get stuck in local minima. Secondly, the input depth data is transformed to match the calibration model and not vice versa. Thirdly, due to the planar geometry of the calibration target, the point-to-plane distance

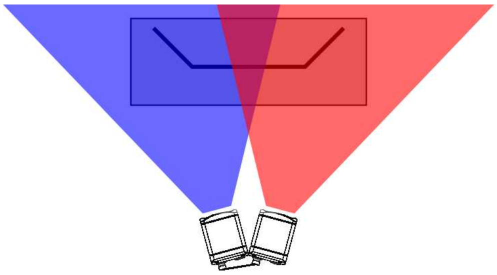

Figure 5: Calibration target as seen by each camera.

will provide a more robust error metric than the standard point-to-point distance. Lastly, since ToF cameras suffer from multi-path interference, edges might be represented inaccurately. Therefore, the edges are eroded in the actual calibration model.

A complete overview of the calibration method is shown in figure 6. An initial estimate of the transformation matrix is obtained by measuring the distance between either camera and the corresponding calibration part, and taking into account the angle between the two cameras. Next, the ICP algorithm is performed to refine these transformation estimates. Since both calibration parts are referenced in a common coordinate system, the rigid transformation from one camera to the other can be derived. The data from the different cameras can now be fused together.

3.3 Data Fusion

The main goal of data fusion is to combine the data from the ToF cameras such that it can be used by an object detection framework. The registered point clouds could simply be summed into a larger point cloud. However, the core of our detection framework is based on 2D detection methods for computational reasons. Therefore, the combined point cloud must be projected onto an image plane. The simple and well-known pinhole camera model is used to describe the 3D to 2D projection. To make sure that the complete point cloud fits on the image plane, appropriate model parameters must be selected.

First of all, a single viewpoint is defined for the virtual camera. This viewpoint is chosen such that the horizontal field of view of the virtual camera matches the combined horizontal field of view of the ToF cameras as can be seen in figure 7. Consequently, the virtual camera is placed at the intersection of the horizontal boundary field of view vectors. However, since the ToF cameras are not perfectly aligned, the viewing rays will not intersect. Therefore the intersection point is calculated as the point that minimizes the distance to each ray.

Next the intrinsic parameters of the virtual camera

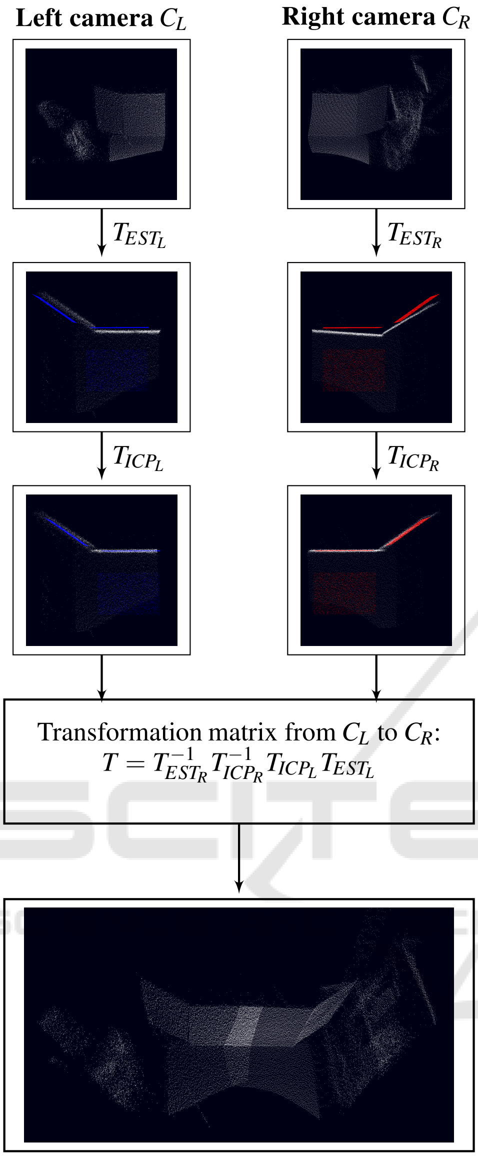
Left camera C_L
Right camera C_R
T_{EST_L}
T_{EST_R}
T_{ICP_L}
T_{ICP_R}
Transformation matrix from C_L to C_R:
$T = T_{EST_R}^{-1} T_{ICP_R}^{-1} T_{ICP_L} T_{EST_L}$

Figure 6: Overview of the calibration method.

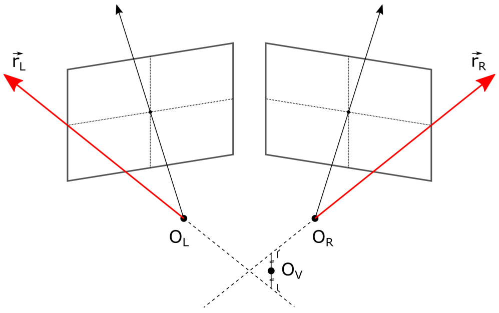

Figure 7: Sensor origin of the virtual camera.

are chosen in accordance with the original ToF cameras. Based on the extrinsic parameters between the two ToF cameras the angle of view can be estimated. The focal length of the ToF camera is maintained. Since the angle of view and focal length are now fixed, the resolution of the image plane is fixed as well. Each point in the point cloud is then projected on the camera image using the pinhole camera model. This way a range image is obtained that can be used for object detection as shown in the next section.

4 RESULTS

4.1 Registration Experiment

To obtain an idea of the accuracy of the calibration, the camera-setup was pointed towards a wall. Using the method described previously, the point clouds were combined into a range image. Next a plane was fitted to the wall. In figure 8, the difference between each point of the wall and the mesh representation of the plane is shown. The color indicates the distance between the fitted plane and wall. When the distance is small, the point is colored green. The histogram has a Gaussian distribution due to the noisy nature of the range data itself. The root mean square error between the fitted plane and wall is $9.84mm$, which corresponds to the relative accuracy of the camera. If there would have been a discrepancy between both due to incorrect calibration, then the histogram would have been more skewed or have outliers.

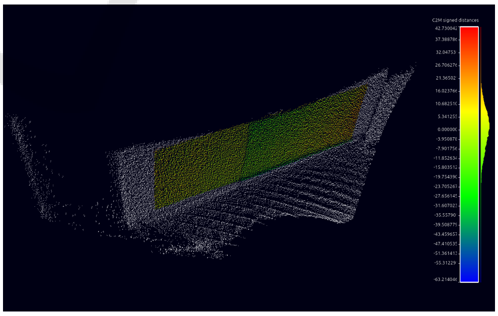

Figure 8: Accuracy of the calibration.

4.2 Application Example

Existing 3D object detection methods can be slow due to the 3D complexity (Abbeloos and Goedemé, 2016). The core idea of our object detection framework is therefore to reduce the 3D problem into a 2D space. The work flow of our object detection framework is shown in figure 9. Firstly, the point clouds are merged using the extrinsic parameters obtained during calibration. Then the combined point cloud is projected onto a range image. Using a template matching technique (e.g. LINEMOD (Hinterstoisser et al., 2012)), the object can be detected in the range image. Lastly, all 2D detection results can be re-projected back into 3D space for an object location refinement. Each 2D detection provides two coordinates, while the third coordinate is estimated from the available depth information. The pose refinement can, for example, be achieved with the ICP algorithm.

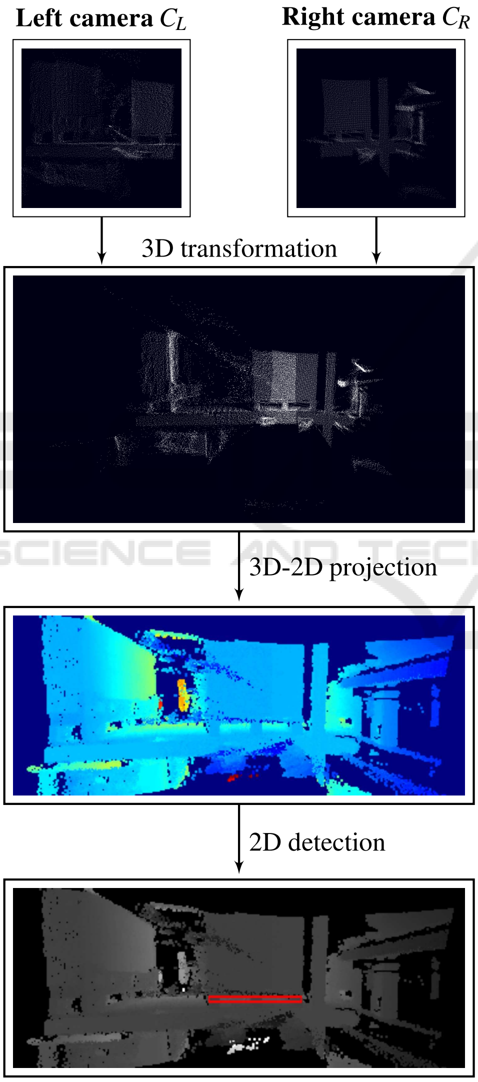

Figure 9: Object detection in an industrial application.

5 CONCLUSIONS

We proposed an accurate and easy-to-use technique for extrinsic calibration of two ToF cameras that are placed side-by-side with only a small overlap between the views. We demonstrated that using only a part of the calibration target in each view, ICP can be used to register both views despite the limited overlap. The calibration target that has been used consists of four planar regions. This has the benefit that it is more robust to noisy range data. Furthermore, for the calibration only one shot of the calibration target is required. The effectiveness of our method was also proven in a real-life application.

REFERENCES

Abbeloos, W. and Goedemé, T. (2016). Point pair feature based object detection for random bin picking. In *Proceedings CRV 2016*, number accepted. University of Victoria.

Auvinet, E., Meunier, J., and Multon, F. (2012). Multiple depth cameras calibration and body volume reconstruction for gait analysis. In *Information Science, Signal Processing and their Applications (ISSPA), 2012 11th International Conference on*, pages 478–483. IEEE.

Besl, P. J. and McKay, N. D. (1992). Method for registration of 3-d shapes. In *Robotics-DL tentative*, pages 586–606. International Society for Optics and Photonics.

Bielicki, J. and Sitnik, R. (2013). A method of 3d object recognition and localization in a cloud of points. *EURASIP Journal on Advances in Signal Processing*, 2013(1):1.

Chetverikov, D., Svirko, D., Stepanov, D., and Krsek, P. (2002). The trimmed iterative closest point algorithm. In *Pattern Recognition, 2002. Proceedings. 16th International Conference on*, volume 3, pages 545–548. IEEE.

Hinterstoisser, S., Cagniart, C., Ilic, S., Sturm, P., Navab, N., Fua, P., and Lepetit, V. (2012). Gradient response maps for real-time detection of textureless objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(5):876–888.

Low, K.-L. (2004). Linear least-squares optimization for point-to-plane icp surface registration. *Chapel Hill, University of North Carolina*, 4.

Ruan, M. and Huber, D. (2014). Calibration of 3d sensors using a spherical target. In *2014 2nd International Conference on 3D Vision*, volume 1, pages 187–193. IEEE.

Zhang, Z. (2000). A flexible new technique for camera calibration. *IEEE Transactions on pattern analysis and machine intelligence*, 22(11):1330–1334.