A Novel 2.5D Pattern for Extrinsic Calibration of ToF and Camera Fusion System

Jiyoung Jung, Yekeun Jeong, Jaesik Park, Hyowon Ha, James Dokyoon Kim, and In-So Kweon

Abstract— **Recently, many researchers have made efforts for accurate calibration of a Time-of-Flight camera to fully utilize its provided depth values. Yet most previous works focus mainly on intrinsic calibration by modeling its systematic errors and noises while extrinsic calibration is also an important factor when constructing sensor fusion system. In this paper, we present a calibration process that can correctly transfer the depth measurements onto the color image. We use 2.5D pattern so that sufficient reprojection error can be considered for both color and ToF cameras. The issues on obtaining the correct correspondences for this pattern are discussed. In the optimization stage, the depth constraint is also employed to ensure the depth measurements to lie on the pattern plane. The strengths of the proposed method over previous approaches are evaluated in several robotic applications which require precise ToF and camera calibration.**

I. INTRODUCTION

Accurate depth estimation of the scene has been one of the key research interests for past decades. This field is essential for a wide spectrum of robot applications, especially regarding navigation related tasks such as path-planning, obstacle avoidance, and 3D mapping. However, image based depth estimation often results in an inaccurate solution due to its inevitable ambiguity. Therefore, the need for metric depth measurement has led people to use such devices as 2D laser range finders and 3D Time-of-Flight cameras.

A. Metric sensors for robots

2D laser range finders are commonly used to a large number of today's mobile robots due to their high speed and accuracy as well as large operational ranges. They are shown to be very effective for various tasks of mobile robots including map building, localization, and obstacle detection[6], [18], [16]. However, since a 2D LRF scans a line at a time, it is usually equipped on a robot platform to scan its surroundings horizontally, and therefore the depth measurement is limited to the horizontal plane at the height of the sensor. Surmann et al.[18] propose a 3D LRF system by mounting a 2D LRF on a standard servo, which is controlled to rotate the mounted sensor vertically. Though it takes a few seconds to grab enough 3D points to represent the scene, the strongest advantage of this system is that it is much cheaper than a 3D Time-of-Flight sensor at that time.

A 3D Time-of-Flight camera modulates its illumination LEDs, and the CCD/CMOS imaging sensor measures the phase and the amplitude of the returned signal at each pixel. As the price of a 3D ToF camera has decreased recently with better accuracy and higher framerate in smaller size, many researchers have installed the sensor on the mobile robots and investigated its feasibilities[20], [15]. Weingarten et al.[20] compare the characteristics of a 2D LRF and a 3D ToF camera carefully and introduce a system that makes use of a 3D ToF camera effectively for navigation of a mobile robot. May et al.[15] compare a highly accurate, expensive 3D laser scanner with a 3D ToF camera which is based on the photon mixer device technology[4]. Presumed that the reliable calibration, lighting adaptation and accuracy filtering is applied, a 3D ToF camera is shown to have several advantages over a 2D LRF and a 3D laser scanner including its small size, weight and mainly its high performance with up to 30 frames per second.

B. Sensor fusion systems and calibration

In a sensor fusion system, a precise calibration of one sensor to another is essential in order to analyze the combined information effectively. There are a number of studies regarding calibration[7], [2], [3]. One of the most popular calibration methods is proposed by Zhang[22]. It is a homography based camera calibration process that uses 2D metric information of the checkerboard plane. This method has been extended to extrinsic calibration of a camera and a 2D laser range finder with the constraint on depth measurements[21].

A 3D ToF camera provides an amplitude image which represents the amplitude of the returned signals as well as the 3D point cloud of the scene. One might think that since the ToF camera provides the amplitude images which are very similar with the traditional grayscale images, the existing calibration methods would work successfully on estimating camera pose between a ToF and a color camera as well. However, the methods dedicated to estimate intrinsic and extrinsic parameters of color cameras[22] or to extrinsic calibration of a camera with a 2D laser range finder[21] do not provide an exact solution to our problem.

The main reason that the traditional calibration methods do not work well on calibrating camera-ToF sensor fusion system is the characteristics of the ToF sensor. Unlike a 2D laser range finder, the depth measurement that a typical 3D ToF camera provides is inaccurate, and the amplitude images

J. Jung, Y. Jeong, J. Park, H. Ha, and I. Kweon are with the Department of Electrical Engineering, KAIST, Daejeon, South Korea {jyjung, ykjeong, jspark, hwha}@rcv.kaist.ac.kr, iskweon@kaist.ac.kr

J. D. Kim is with Samsung Advanced Institute of Technology, South Korea jamesdk.kim@samsung.com

This work is supported by supported by National Strategic R&D Program for Industrial Technology, Korea. This work is also supported by the Defense Acquisition Program Administration and the Agency for Defense Development, Korea through the Image Information Research Center at Korea Advanced Institute of Science and Technology under the contract UD100006CD.

are in low resolution compared with general color images. Therefore the process of calibrating sensor fusion system which includes a ToF camera has to be designed to overcome the weaknesses of the sensor.

Kim et al.[10] and Fuchs and Hirzinger[5] focus on 3D ToF sensor fusion system but they concentrate most of their efforts on intrinsic calibration by modeling the systematic bias, errors and noises of the sensor. Kahlmann et al.[8] and Lindner and Kolb [13] use the calibration method of [22] and [3] on the amplitude images to estimate internal and external parameters. The depth measurement is not included for pose estimation and therefore they do not overcome the low resolution of the amplitude images. Schiller et al.[17] adds the depth measurement constraint to the method using a checkerboard plane, but due to the low resolution of the amplitude images, manual correspondence selection is inevitable. Kern et al.[9] uses a plane with holes but their objective is to calibrate a laser scanner which provides much more accurate depth measurements compared with a 3D ToF camera. In addition, since their holes are arranged in a grid, they have to go through another algorithm to identify the holes, whereas our pattern has holes spread uniquely so that identification process becomes very simple.

In this paper, we present an extrinsic calibration process of the sensor fusion system which consists of a color camera and a 3D Time-of-Flight camera. We have designed a novel 2.5D pattern with 4cm-diameter holes so that the correct correspondences are obtained automatically. The initial estimates of the intrinsic and extrinsic parameters of both cameras with respect to the pattern plane are obtained using the homography based calibration[22] implemented in the open source library[19]. The obtained intrinsic parameters of the ToF camera are highly incorrect due to the severe radial distortion of the amplitude images. We refine the intrinsic parameters and remove the radial distortion from the images. The camera motion of the color camera with respect to the pattern plane and that of the ToF camera with respect to the color camera are optimized by minimizing reprojection errors and depth measurement errors. The process is carefully planned to include the considerations of the correct correspondence acquisition for reprojection error minimization, the constraint on the depth measurements, and the appropriate pinhole camera modeling of a ToF sensor.

The rest of the paper is organized as follows: Section II explains the traditional calibration methods which are popularly used as well as their limitations. Section III explains our approach with details. The effectiveness of our method is shown in Section IV by performing demonstrations that require a precise extrinsic calibration of the sensor fusion system. Section V concludes this paper.

II. CAMERA AND LASER CALIBRATION METHODS

A. Homography based camera calibration

The homography based camera calibration requires the camera to observe a particular planar pattern shown at a few different orientations. The pattern is usually checker so that it is easy to obtain correspondences using corner detector. The homography between the model plane and its image is calculated to get the constraints on the intrinsic parameters.

Let $\mathbf{m} = [u, v, 1]^T$ and $\mathbf{M} = [X, Y, Z, 1]^T$ be a 2D and a 3D point in homogeneous representation respectively. In a pinhole camera model, the relationship between a 3D point $\mathbf{M}$ and its image projection $\mathbf{m}$ is given by

$$s\mathbf{m} = \mathbf{K}[\mathbf{R} \quad \mathbf{t}]\mathbf{M} \tag{1}$$

where s is an arbitrary scale factor. $\mathbf{R}$ and $\mathbf{t}$ are the rotation matrix and the translation vector which relate the world coordinate system to the camera coordinate system respectively, and $\mathbf{K}$ is the camera intrinsic matrix. Assuming the model plane is on $Z = 0$ of the world coordinate system, we have

$$s[u, v, 1]^T = \mathbf{K}[\mathbf{r_1} \quad \mathbf{r_2} \quad \mathbf{t}][X, Y, 1]^T \tag{2}$$

where $\mathbf{r_i}$ is the i^{th} column of the rotation matrix $\mathbf{R}$. Given an image of the model plane, the homography $\mathbf{H} = [\mathbf{h_1} \quad \mathbf{h_2} \quad \mathbf{h_3}]$ can be estimated so that we have

$$[\mathbf{h_1} \quad \mathbf{h_2} \quad \mathbf{h_3}] = \lambda\mathbf{K}[\mathbf{r_1} \quad \mathbf{r_2} \quad \mathbf{t}] \tag{3}$$

From (3), we obtain the following constraints on the camera intrinsic matrix $\mathbf{K}$ using the fact that $\mathbf{r_1}$ and $\mathbf{r_2}$ are orthonormal.

$$\begin{aligned} \mathbf{h_1^T K^{-T} K^{-1} h_2} &= 0 \\ \mathbf{h_1^T K^{-T} K^{-1} h_1} &= \mathbf{h_2^T K^{-T} K^{-1} h_2} \end{aligned} \tag{4}$$

These constraints construct a homogeneous equation in $\mathbf{Ax} = \mathbf{0}$ form, which is solvable using singular value decomposition when sufficient images are provided. Once $\mathbf{K}$ is known, the extrinsic parameters for each image are easily computed. From (3), we have

$$\begin{aligned} \mathbf{r_1} &= \lambda\mathbf{K^{-1}h_1} \\ \mathbf{r_2} &= \lambda\mathbf{K^{-1}h_2} \\ \mathbf{r_3} &= \mathbf{r_1} \times \mathbf{r_2} \\ \mathbf{t} &= \lambda\mathbf{K^{-1}h_3} \end{aligned} \tag{5}$$

with $\lambda = 1/\left\|\mathbf{K^{-1}h_1}\right\| = 1/\left\|\mathbf{K^{-1}h_2}\right\|$. The solution is further refined through maximum likelihood inference and the radial distortion of the camera lens is considered in the final nonlinear minimization.

In order to obtain an accurate calibration result from this method, the correct correspondences between the model plane and its images are crucial. The number of corners has to be sufficient to calculate homography and they have to be widely spread to each boundary of the images to get the reliable estimation of the radial distortion parameters.

B. Camera-laser calibration

There has been a number of studies regarding the pose estimation of a camera with respect to a 2D laser range finder[1][21]. The key idea is to take advantage of the depth measurements from the laser scanner and restrict them onto the calibration plane in addition to the homography based camera calibration.

In [21], a camera-laser scanner rig is considered. The intrinsic and extrinsic parameters of the camera with respect to the checkerboard are estimated using the homography based method. The transformation of the point $\mathbf{M}$ in the camera coordinate system to the point $\mathbf{Q}$ in the laser coordinate system can be represented as,

$$\mathbf{Q} = [\mathbf{R_L} \quad \mathbf{t_L}]\mathbf{M} \tag{6}$$

where $\mathbf{R_L}$ and $\mathbf{t_L}$ are the rotation matrix and the translation vector of the camera relative to the laser range finder respectively. In order to estimate $\mathbf{R_L}$ and $\mathbf{t_L}$, the checkerboard plane is parametrized by a 3-vector $\mathbf{N}$ such that $\mathbf{N}$ is parallel to the normal of the plane and its magnitude, $\left\|\mathbf{N}\right\|$, equals to the distance from the camera to the plane. From (1),

$$\mathbf{N} = -\mathbf{r_3}(\mathbf{r_3^T t}) \tag{7}$$

The laser point $\mathbf{Q}$ in the camera coordinate system described as $\widetilde{\mathbf{M}} = \mathbf{R_L^{-1}}(\mathbf{Q} - \mathbf{t_L})$ must satisfy $\mathbf{N} \cdot \widetilde{\mathbf{M}} = \left\|\mathbf{N}\right\|^2$ to lie on the checkerboard plane. Then we have

$$\mathbf{N} \cdot \mathbf{R_L^{-1}}(\mathbf{Q} - \mathbf{t_L}) = \left\|\mathbf{N}\right\|^2 \tag{8}$$

For the measured calibration plane parameter $\mathbf{N}$ and the laser point $\mathbf{Q}$, this gives the depth constraint on $\mathbf{R_L}$ and $\mathbf{t_L}$.

Since the plane parameters are computed using the camera motions, this process depends heavily on good initial estimates of the camera motions. For a color camera, the homography based method using the popular checker pattern gives reasonable intrinsic and extrinsic calibration results because a set of correct correspondences can be secured. However, in case of a 3D Time-of-Flight camera, the resolution of an amplitude image is very low (e.g. 176x144, MESA^{TM} Swissrangeer 4000) relative to the resolution of a general color image (e.g. 1280x960, Pointgrey Flea2) and the radial distortion is severe so that it is hard to obtain the correspondences using a general corner detector on checkerboard images.

Additionally, the depth constraint is based on the assumption that the depth measurements from a laser range finder and the corresponding ray direction is trustworthy. Unlike a 2D laser range finder, a 3D ToF camera contains several LEDs aligned in two-dimension. The interference between the different infrared rays affects the accuracy of the depth measurement. Therefore, although it is possible to modify the depth constraint on a 2D laser range finder to be applied to the 3D measurements of a ToF camera, there has to be a solution to overcome the inaccurate depth measurements of the ToF camera in the extrinsic calibration process of camera-ToF sensor fusion system. In our method, a similar depth constraint is applied in the optimization process with preprocessing and modification.

III. THE PROPOSED CALIBRATION

The proposed method makes use of the homography based calibration. However, we present the solutions to deal with the problems mentioned above, such as radially distorted low resolution images and inaccurate depth measurements of a ToF camera. We take the depth measurements into account in the optimization process as well as the reprojection errors computed from the correspondences between the specifically designed calibration plane and the images.

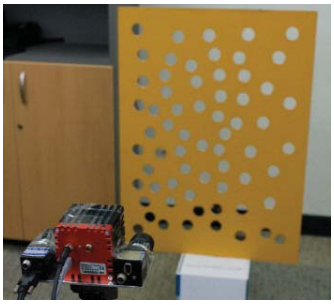


Fig. 1. The 2.5D pattern plane and the camera-ToF sensor fusion system.

TABLE I
DETECTED CORRESPONDENCES IN THE TOF AMPLITUDE IMAGES WITH AND WITHOUT THE RADIAL DISTORTION

	Color images	ToF amplitude images
Radially distorted	2481	1287
Radially undistorted	2481	2087

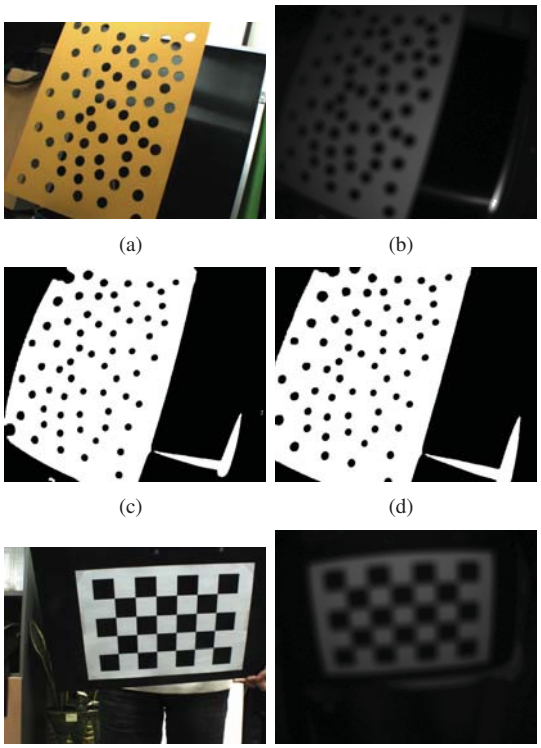
A. 2.5D pattern plane

To attain a large number of precise correspondences between the model plane and its images for the calibration of a ToF camera, we need a new type of pattern that can assure a reliable feature detection in a blurry low resolution image. Inspired by [19], we use the pattern consists of black holes irregularly placed on a plane. They are more robust to be detected in a low resolution image because the center of a circular pattern is preserved when the image is isotropically blurred. Moreover, the matching(i.e. the identification) process of the detected circles in the image to the pattern plane becomes simple because the dots are irregularly spread.

In order to obtain an amplitude image from a ToF camera as clear as the color image of a black-and-white checker pattern, we have not printed the pattern on a plane. Instead, we have particularly designed a 2.5D pattern plane of which the features can be recognized by the distinctive differences in the amplitude of the received rays caused by the depth variations as well as the differences in the color of the plane.

Since the goal is to find the correspondences of the pattern plane with the amplitude images of a ToF camera as well as with the images of a color camera, we have constructed a 80cmx60cm pattern board that has 64 holes, as shown in Figure 1. The diameter of a hole is 4cm, which is large enough for the infrared rays to pass through the hole so that the circular patterns are clearly shown in a 176x144-sized amplitude image.

In Figure 2-(a) the 2.5D pattern plane is shown in a 1280x960 resolution color image. Without loss of generality, we assume that the pattern plane is at $Z = 0$ in the world coordinate system and the coordinates of the centers of the holes are known. We have designed some careful preprocessing steps that guarantee more correspondences from the ToF images. Since the ellipses are still difficult to



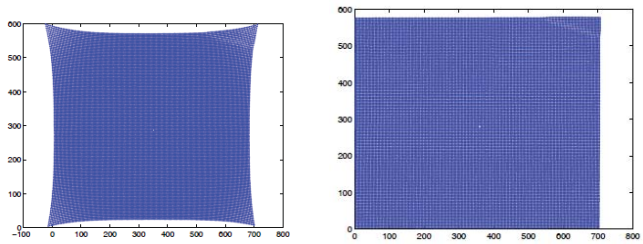
(a) (b) (c) (d) (e) (f)

Fig. 2. The 2.5D dot pattern plane (a) in a 1280x960 color image, (b) in a 176x144 ToF amplitude image. (c) The binary image of (b), and (d) the radially undistorted binary image. (e) A traditional checkerboard plane in a 1280x960 color image, and (f) in a 176x144 ToF amplitude image.

be detected in the original low resolution amplitude images (Figure 2-(b)), we enlarge the image by 4 times using the bicubic interpolation and transform them into the binary images using the constant thresholds (Figure 2-(c)). It is shown that the image coordinates of the center of the holes are still well preserved.

However, since the detected ellipses from the image are considered as the actual holes on the pattern only if they satisfy the homography constraint, many correctly detected ellipses are removed from correspondences, considered as false detection because of the severe radial distortion. Therefore we use radially undistorted images (Figure 2-(d)) so that more holes in the image satisfy the homography constraint and to be considered as the correct correspondences. Table 1 shows the total number of detected correspondences using [19]. We have used 47 color images and 47 ToF amplitude images in the calibration process. By removing radial distortions from the amplitude images, the number of correspondences are increased significantly yielding better calibration results. This process is fully automatic so that no manual feature selection is needed.

Optimizing the intrinsic parameters for precise distortion removal is explained in the next section. In a high resolution color image and a low resolution ToF amplitude image of a traditional checkerboard are shown in Figures 2-(e) and (f) for comparison. The ToF image is too blurry to detect corners for acquisition of the correct correspondences.



(a) (b)

Fig. 3. The projected 3D measurements of a ToF camera (a) using the initial estimates of the intrinsic parameters obtained by the homography based method, and (b) using the intrinsic parameters refined by the pinhole camera model.

B. Intrinsic parameters of the ToF camera

A ToF camera provides the depth measurements for every pixel. So the relation between the 3D points in the ToF camera coordinate system and their projections are known by the manufacturer. In this work, we do not estimate the systematic errors or noises to improve the intrinsic calibration of the ToF camera and we rather focus on the extrinsic calibration between a ToF camera and a color camera.

However, we need the intrinsic parameters for the pinhole camera model in order to remove the radial distortions from the amplitude images and to calculate the reprojection errors in the optimization process. The homography based calibration using the correspondences between the 2.5D pattern plane and the amplitude images with radial distortion gives us the initial estimates of the intrinsic parameters of the ToF camera. Figure 3-(a) shows the reprojected 3D measurements of ToF camera using the initial estimates of the intrinsic parameters. The projected points are more distorted as the image coordinate goes farther from the center due to the incorrect intrinsic parameters. It is because the intrinsic parameters are calculated with insufficient circular features mostly around the center of the amplitude images, instead of considering every depth measurement of an image which is essential for calibrating the ray directions.

Therefore we further refine the intrinsic parameters so that the manufacturer-provided relation between the depth measurements and their projections can be modeled as a pinhole camera model. We use LM optimization to minimize the reprojection errors between all 176x144 depth measurements and their corresponding locations on the image coordinate. The projected points in Figure 3-(b) are shown to be rectangular, which means the refined intrinsic parameters successfully model the ToF camera projections, i.e., the optimized intrinsic calibration successfully maps the 3D points onto the amplitude image coordinate.

C. LM optimization

Once the correspondences of the pattern plane in the color images and in the ToF amplitude images are obtained, and the correct intrinsic parameters of both the color and ToF cameras are known, we can find the optimal pose of the ToF camera with respect to the color camera by minimizing the reprojection errors. Let the feature point in the i^{th} color

image be $\mathbf{x_{ci}}$ and in the i^{th} ToF amplitude image be $\mathbf{x_{ti}}$. The image projection of the 3D point $\mathbf{X_{plane}}$ onto the i^{th} color image is $\widetilde{\mathbf{x_{ci}}}$, the i^{th} ToF amplitude image is $\widetilde{\mathbf{x_{ti}}}$;

$$\widetilde{\mathbf{x_{ci}}} = \mathbf{K_c}[\mathbf{R_{ci}} \quad \mathbf{t_{ci}}]\mathbf{X_{plane}} \tag{9}$$

$$\widetilde{\mathbf{x_{ti}}} = \mathbf{K_t}[\mathbf{R_{c2t}} \quad \mathbf{t_{c2t}}] \begin{bmatrix} \mathbf{R_{ci}} & \mathbf{t_{ci}} \\ \mathbf{0} & 1 \end{bmatrix} \mathbf{X_{plane}} \tag{10}$$

where $\mathbf{K_c}$ and $\mathbf{K_t}$ are the intrinsic camera matrices of the color camera and the ToF camera respectively. $[\mathbf{R_{ci}} \quad \mathbf{t_{ci}}]$ is the i^{th} color camera pose with respect to the world coordinate system and $[\mathbf{R_{c2t}} \quad \mathbf{t_{c2t}}]$ is the ToF camera pose with respect to the color camera coordinate system. Then, the problem can be represented as an error minimization,

$$minf(\mathbf{P}) = min \sum_i (\left\|\mathbf{x_{ci}} - \widetilde{\mathbf{x_{ci}}}\right\|^2 + w_t \left\|\mathbf{x_{ti}} - \widetilde{\mathbf{x_{ti}}}\right\|^2) \tag{11}$$

where $\mathbf{P}$ represents the extrinsic camera parameters of $[\mathbf{R_{c2t}} \quad \mathbf{t_{c2t}}]$ and $[\mathbf{R_{c_i}} \quad \mathbf{t_{c_i}}]$ for all i's. Because the 3D point $\mathbf{X_{plane}}$ are known and fixed, this optimization goes to more like a camera-resectioning with a constraint that those two sensors are relatively fixed.

Using the camera motions given by the homography based calibration as the initial estimates, the reprojection error is minimized using Levenberg-Marquardt optimization technique[12], [14]. The optimized camera motions are then used to estimate the plane parameters of the pattern plane in the similar way described in Section II-B in order to enforce the depth constraint. To constrain the depth measurements of the plane to be precise, we first distinguish the depth measurements that should lie on the pattern plane by filtering out wrongful depth values at the plane boundaries and around the holes using the plane model prior as shown in Figure 4. Figure 4-(a) are the depth measurements which fall within constant thresholds to XYZ coordinates. It is shown that there are some outliers at the bottom boundary of the pattern plane which do not belong to the plane. These outliers are removed by estimating the plane parameters using singular value decomposition (Figure 4-(b)).

Including the depth constraint, which restricts the selected depth measurements to be on the plane $Z = 0$, the minimization problem alters as follows.

$$minf(\mathbf{P}) = min(\sum_i (\left\|\mathbf{x_{ci}} - \widetilde{\mathbf{x_{ci}}}\right\|^2 + w_t \left\|\mathbf{x_{ti}} - \widetilde{\mathbf{x_{ti}}}\right\|^2) \ldots + \sum (w_p \mathbf{X_p}(\mathbf{Z}))) \tag{12}$$

w_t and w_p are the weights to balance the three different errors. In our experiments, $w_t = 1/4$ and $w_p = 1/100$ which mean that 4-pixel error in the ToF reprojection and 100mm error in the depth measurements are considered the same as 1-pixel error in the color camera reprojection. 1 to 4 ratio of the color-ToF camera reprojection error is reasonable because the ToF images are upsampled by 4 times in the experiments for better correspondence search.

Table 2 shows the effectiveness of each step in our method. We have used 47 color images and 47 ToF amplitude images with radial distortion removed. We have experimented with

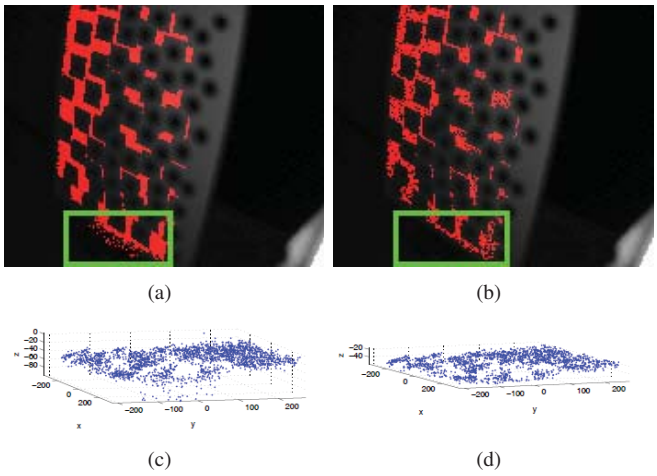


Fig. 4. Filtering out depth measurements to enforce the depth constraint (a) by applying constant thresholds and (b) by applying plane constraint additionally. The outliers at the bottom boundary of the pattern plane are removed. (c) and (d) are the depth measurements in 3D space, which are bottom side views of (a) and (b) respectively.

TABLE II

AVERAGE REPROJECTION ERRORS [PIX] OF COLOR AND TOF CAMERA AND DEPTH MEASUREMENT ERRORS [MM]

Intrinsic Refine.	Repr.Err. Min.	Depth Constraint	Color Repr.Err.	ToF Repr.Err.	Depth Error
1) X	X	X	0.237	2.276	20.419
2) O	X	X	0.237	2.247	17.456
3) X	O	X	0.789	1.632	19.265
4) O	O	X	0.777	1.630	16.737
5) O	O	O	0.804	1.666	8.194

5 different settings: with or without intrinsic refinement, with or without reprojection error minimization on color and ToF images, and adding depth measurement constraint. Comparing the experiment 1 with 2 and 3 with 4, the depth measurement errors are decreased because the provided 3D measurements are aligned with their locations on the ToF images in the intrinsic refinement process. Comparing the experiment 2 with 4, the reprojection errors of ToF images are decreased at the cost of the increased errors of color images. However, both reprojection errors are still 1-2 pixels which shows the effectiveness of the reprojection error minimization using LM optimization. The experiment 5 includes the depth constraint, which decreases the depth measurement errors significantly at the cost of slight increase of reprojection errors.

IV. EXPERIMENTAL RESULTS

We have tested our method and two other previous approaches, namely homography based calibration and camera-laser calibration methods to our sensor fusion system. The system consists of a color camera (Pointgrey Flea2) and a Time-of-Flight sensor (MESA^{TM} Swissranger 4000) as shown in Figure 1. Since it is difficult to get the ground truth extrinsic camera parameters of the two sensors for evaluation, we show the performance of our method by two different applications that require precise extrinsic calibration: 3D rendering and depthmap upsampling[11].

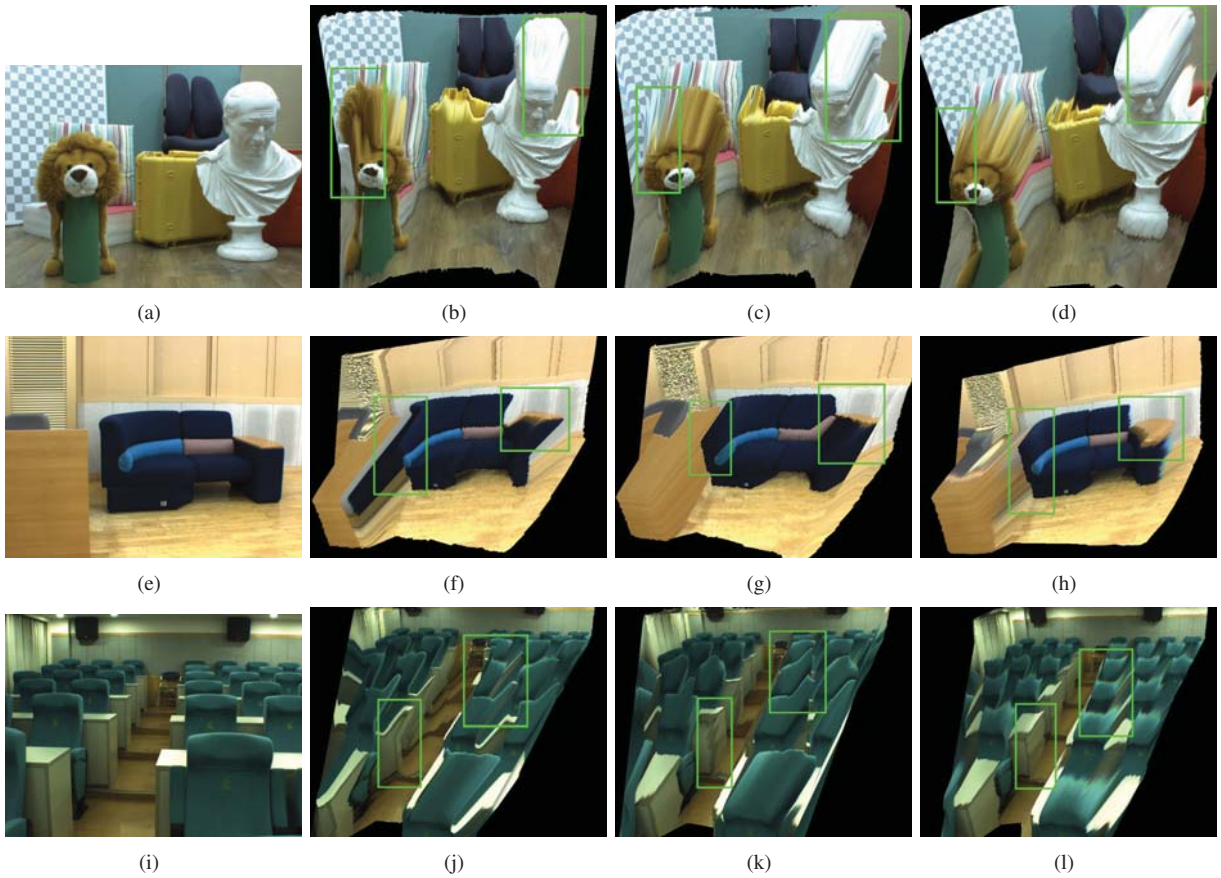


Fig. 5. 3D rendering results after projection of depth measurements onto color images. (a,e,i) original color images. The projection is due to calibration results by (b,f,j) homography based calibration method, (c,g,k) camera-laser calibration method, and (d,h,l) the proposed method.

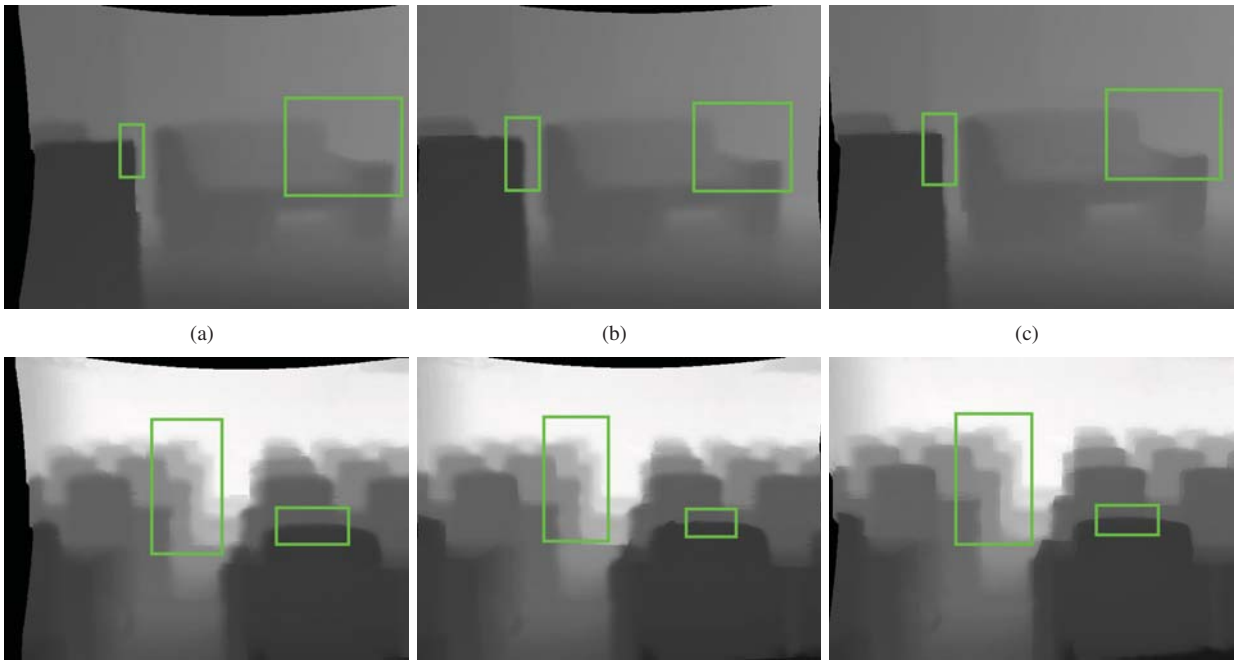


Fig. 6. Depthmap upsampling results using joint bilateral upsampling after projection of depth measurements onto color images. The projection is due to calibration results by (a,d) homography based calibration method, (b,e) camera-laser calibration method, and (c,f) the proposed method.

In Figure 5, we have transfered the depth measurements onto the color image and rendered in 3D space using Virtual Reality Modeling Language (VRML). The transferal of the depth measurements from the ToF camera to the color image has been performed using the extrinsic calibration results by the three different methods. From left to right, each column represents the color image, 3D rendering results using the homography based calibration, the camera-laser calibration, and the proposed method. The dataset includes three different real scenes of various scene depths for evaluation. Figure 5-(a) contains a close scene of which the depth range falls within 2 meters. The depth ranges of the scenes in Figures 5-(e) and (i) are around 4 meters, and 10 meters respectively. The maximum depth measurement of the ToF sensor has been examined as 10 meters.

In the second column, it is shown that the homography based methods completely fails to align the depth measurements onto the color image. The camera-laser calibration shows somewhat better results but still, the depth discontinuities often fails to coincide with object boundaries. The proposed method shown in the rightmost column successfully aligns the depth measurements onto the color image. This means the estimated pose of the camera to the ToF sensor is very much precise.

Another application that we have tested is depthmap upsampling. The ToF camera provides the depthmap of the scene in 176x144 resolution, which is very low compared with the resolution of the color image (1280x960). A simple bicubic interpolation blurs the depthmap at the discontinuities which would eventually lead to problems in further applications due to wrongful depth assignments. Therefore we have performed the joint bilateral upsampling[11] which uses the color guidance of the high-resolution image to upsample the depth measurements. The correct calibration of the two sensors is essential to this technique because the depth measurements have to be assigned in the right location on the color image to be upsampled.

Figure 6 shows the depthmap upsampling results in two different datasets. Each row contains the upsampled depthmaps of the scene in Figure 5-(e) and (i) respectively. For each column from left to right, we have used the homography based calibration, camera-laser calibration, and the proposed method. The upsampled depthmap using the calibration information by the proposed method has clearer boundaries due to correct initial depth assignment. We suggest you to enlarge the figures to check the clearity of the object boundaries.

V. CONCLUSION

In this paper, we have presented an extrinsic calibration method to estimate the pose of a color camera with respect to a 3D Time-of-Flight camera. We use 2.5D pattern so that the correct correspondences are obtained for both color and ToF cameras. For accurate reprojection error calculation, we refine the intrinsic parameters of the ToF camera to model its projection as a pinhole camera model. Depth constraint which restricts the depth measurement to lie on the pattern plane is also employed into LM optimization as well as the reprojection errors. Our process is basically fully automatic, including the acquisition of sufficient correct correspondences. The performance of our method is shown to be very effective for further applications such as 3D rendering and depthmap upsampling, which require exact extrinsic calibration of the camera and the ToF sensor.

REFERENCES

[1] Y. Bok, D. Choi, Y. Jeong, and I. S. Kweon. Capturing village-level heritages with a hand-held camera-laser fusion sensor. In *IEEE Workshop on eHeritage and Digital Art Preservation in conjunction with ICCV 2009*, pages 947 –954, Oct. 2009.

[2] S. Bougnoux. From projective to euclidean space under any practical situation, a criticism of self-calibration. In *the Sixth International Conference on Computer Vision*, pages 790 –796, Jan. 1998.

[3] J. Bouguet. Visual methods for three-dimensional modelling. *PhD. thesis*, 1999.

[4] C. C. S. dElectronique et de Microtechnique SA. Swiss ranger sr-2. 2005.

[5] S. Fuchs and G. Hirzinger. Extrinsic and depth calibration of tof-cameras. In *IEEE Conference on Computer Vision and Pattern Recognition,*, pages 1 –6, 2008.

[6] J. Gonzalez, a. Ollero, and a. Reina. Map building for a mobile robot equipped with a 2d laser rangefinder. In *Proceedings of the 1994 IEEE International Conference on Robotics and Automation*, pages 1904–1909. IEEE Comput. Soc. Press, 1994.

[7] R. Hartley. An algorithm for self calibration from several views. In *Proceedings of CVPR 1994*, pages 908 –912, June 1994.

[8] T. Kahlmann, F. Remondino, and H. Ingensand. Calibration for increased accuracy of the range imaging camera swissrangerTM. In *ISPRS Image Engineering and Vision Metrology*, 2006.

[9] F. Kern. Supplementing laser scanner geometric data with photogrammetric images for modeling. *18th International CIPA Symposium, Potsdam*, pages 1–8, 2001.

[10] Y. M. Kim, D. Chan, C. Theobalt, and S. Thrun. Design and calibration of a multi-view tof sensor fusion system. In *Computer Vision and Pattern Recognition Workshops, CVPRW. IEEE Computer Society Conference on*, pages 1 –7, 2008.

[11] J. Kopf, M. F. Cohen, D. Lischinski, and M. Uyttendaele. Joint bilateral upsampling. *ACM Trans. Graph.*, 26, July 2007.

[12] K. Levenberg. A method for the solution of certain problems in least squares. *Quart. Appl. Math.*, 2, 1944.

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

[14] D. Marquardt. An algorithm for least-squares estimation of nonlinear parameters. *SIAM J. Appl. Math.*, 11, 1963.

[15] S. May, B. Werner, H. Surmann, and K. Pervolz. 3D time-of-flight cameras for mobile robotics. *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 790–795, Oct. 2006.

[16] V. Nguyen, A. Martinelli, N. Tomatis, and R. Siegwart. A comparison of line extraction algorithms using 2D laser rangefinder for indoor mobile robotics. In *IROS 2005*, pages 1929–1934. IEEE, 2005.

[17] I. Schiller, C. Beder, and R. Koch. Calibration of a PMD-camera using a planar calibration pattern together with a multi-camera setup. In *ISPRS Congress*, page B3a: 297 ff, 2008.

[18] H. Surmann, K. Lingemann, A. N, J. Hertzberg, and S. Augustin. A 3D laser range finder for autonomous mobile robots. *Symposium A Quarterly Journal In Modern Foreign Literatures*, (April):153 – 158, 2001.

[19] G. Vogiatzis and C. Hernndez. Automatic camera pose estimation from dot pattern. *http://george-vogiatzis.org/calib/*, 2010.

[20] J. Weingarten, G. Gruener, and R. Siegwart. A state-of-the-art 3D sensor for robot navigation. In *Proceeding of 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems*, volume 3, pages 2155–2160, 2004.

[21] Q. Zhang and R. Pless. Extrinsic calibration of a camera and laser range finder (improves camera calibration). In *Intelligent Robots and Systems (IROS), IEEE/RSJ International Conference on*, 2004.

[22] Z. Zhang. A flexible new technique for camera calibration. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(11):1330 – 1334, Nov. 2000.