https://doi.org/10.2352/ISSN.2470-1173.2019.4.PMII-353

Self-calibrated Surface Acquisition for Integrated Positioning Verification in Medical Applications

Sven Jörissen; Informatics VII – Robotics and Telematics, Julius Maximilian University of Würzburg; Germany
Michael Bleier; Zentrum für Telematik e.V., Würzburg; Germany
Andreas Nüchter; Informatics VII – Robotics and Telematics, Julius Maximilian University of Würzburg; Germany

Abstract

This paper presents a novel approach for a position verification system in medical applications. By replacing the already existing cross line laser projectors with galvo- or MEMS-based projectors and utilizing the surveillance cameras, a self-calibration of the system is performed and surface acquisition for positioning verification is demonstrated. The functionality is shown by analyzing the radii of calibration spheres and determining the quality of the captured surface with respect to a reference model. The paper focuses on the demonstration with one pair of camera and projector but can also be extended to a multi-camera-projector system, as present in treatment rooms. Compared to other systems, this approach does not need external hardware and is thus space and cost efficient.

Introduction

Nowadays, a wide range of medical applications demand accurate patient positioning for a successful treatment. While the positioning for X-ray imaging allows tolerances of several millimeters since typically rather big areas are imaged, the required accuracy for CT-imaging and especially classical radiation therapy, Volumetric Arc Therapy (VMAT), Intensity-Modulated Radiation Therapy (IMRT) and 3D Conformal Radiation Therapy (3D CRT) for cancer treatment is much higher. The goal of radiation therapy is to damage the cancer cells as much as possible, while keeping the amount of radiation within the surrounding tissue to an absolute minimum. The de facto standard procedure for patient positioning in radiation therapy is as follows: An initial CT scan is performed to gather anatomical data for the treatment. Here, markers are placed on the patients skin, which are later used to align the patient with the orthogonal line lasers in the treatment room. In a previous step, those line lasers are calibrated to directly intersect in the linear accelerators (linacs) ”isocenter”, the point where the beams of the rotating linac intercept and therefore the radiation intensity is at its peak. The calibration of the isocenter is done performing the Winston-Lutz test. Once the isocenter is calibrated and the patient aligned, the treatment is started. Typically, the initial CT scans outcome is used for several radiation therapy sessions, so are the markers. Fig. 1 shows a typical treatment room with patient couch, gantry, red room lasers for positioning and a test phantom. The importance of precise patient positioning and the potential of optical surface imaging technologies for both positioning and respiratory gating is becoming more and more clear and was recently confirmed and discussed by publications such as [1], [2] and [3].

This paper provides a new method of verifying the patients position with respect to the linacs isocenter. A typical treatment room already consists of multiple cameras for surveillance and line lasers for calibration, isocenter visualization and patient positioning. By replacing those static line lasers with galvo- or MEMS-based laser projectors and combining the laser projectors with the cameras to active stereo systems, multiple applications are imaginable, while the main functionality of positioning the patient manually with respect to the laser crosses and thus, the isocenter, is still provided:

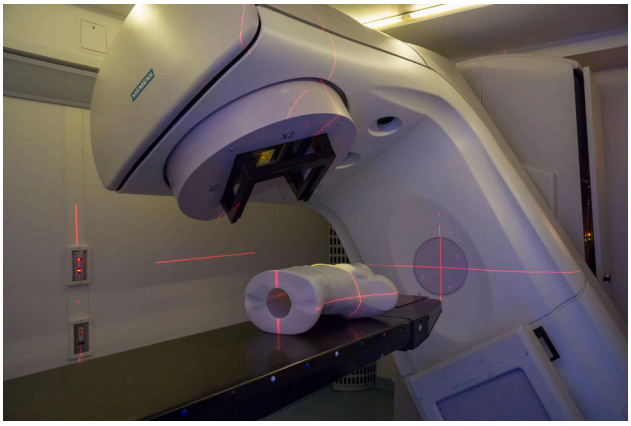

Figure 1. *Radiation-therapy room with gantry and positioning lasers (red)*

1. The extrinsic calibration of the system is performed automatically, making the handling easy and self-verifying.
2. The patient’s position is acquired by scanning the surface using a shape reconstruction method based on light sectioning.
3. The position can then be matched to the data from the CT-scan, giving a translation vector for shifting the position of the patient by adjusting the treatment bench.
4. Additionally, respiratory gating can be performed to increase the efficiency of the therapy and thus, protecting the surrounding tissue during treatment.

Since no CT-scan data or with the Winston-Lutz test aligned set of projectors was present and respiratory gating is still considered to be future work, this paper focuses on the first two applications and demonstrates the functionality by means of a test setup consisting of one camera-projector-pair.

Although the proposed self-calibrating projector-camera setup is applicable for shape acquisition in general when used as a laser-line based structured light system, this paper only focuses

on surface reconstruction in medical applications and a possible combination with respiratory gating. While the previously described manual alignment of patient and isocenter is still widely used, there are already products commercially available, which offer surface guidance prior to and during radiation therapy. The AlignRT system by visionrt [4] uses proprietary 3D stereo cameras and a speckle projector (similar to the Microsoft Kinect) in one housing to recover shape and track the surface to automatically stop the radiation, when the patients motion exceeds certain thresholds. Two of those housings are mounted beneath the wall at a specified angle or are directly attached to the gantry to ensure that the patient is always monitored, even when the gantry is moving. Surface acquisition and position is performed with sub millimeter accuracy. Catalyst HD by the Swedish company C-RAD [5] is another surface guided radiation therapy solution for patient tracking before and during treatment. Consisting of three housings, each with one camera and one projector, mounted on the wall to ensure complete coverage, stereo imaging using triangulation and a non-rigid algorithm are used to monitor the patient and project a color map into the scene to call the attention on possible misplacements. Again, sub millimeter accuracy is guaranteed.

A laser-based system is the Anzai Medical AZ-733VI [6], which is an extension to the Anzai Sensor Belt, a pneumatic solution for respiratory gating. A single point is projected onto the skin of the patient and tracked to account for breathing. While Anzais pressure-sensitive belt is one of the commonly used tools for gating, it is not clear if the laser-based system is actively used in medical treatment for cancer.

While especially AlignRT and Catalyst HD provide good guidance and assistance throughout the whole treatment, all systems require additional hardware and enough space to mount them. This is where the proposed solution in this paper comes into play, since it does not need additional hardware or mounting space but works with slightly modified cross laser projectors and the already available cameras. Although both systems use a principle similar to our approach, different utilizations of stereo are present. While AlignRT uses dual-camera based stereo and a speckle projector for surface reconstruction and respiratory gating, Catalyst HD uses an active stereo system consisting of only one camera and a projector. This is more similar to our camera-laser projector-based system, while still having the need of additional hardware and mounting space, since the traditional cross lasers mentioned earlier are still needed.

Since one of the key aspects of the proposed system is the fact that it is self-calibrating, a short overview on existing self-calibration techniques for structured light systems is given in the following. Calibration of structured light systems is essential for an accurate 3D reconstruction of the scene. The process typically involves either separate or combined calibration of intrinsic parameters for both camera and projector as well as extrinsic parameters. Those procedures are typically long and need to be repeated every time the physical configuration of the system is changed. A self-calibration process for structured light systems was proposed by Furukawa and Kawasaki [7]. The setup consists of a calibrated camera and an uncalibrated video projector and uses uncalibrated stereo for the reconstruction. An additional laser pointer is attached to the projector to determine the scaling parameters. Correspondences between the projection and the image are determined for multiple positions of either a moving camera or a moving projector, resulting in a multi-image 3D reconstruction.

Another method proposed by Aliga and Xu uses uncalibrated projectors and cameras to generate a multi-view 3D point cloud based on a photo-geometric approach [8]. First, an uncalibrated photometric stereo procedure using the projectors as a diffuse light source is performed, then a geometric modelling using the previously estimated surface, approximate lightning directions and reprojection equations is applied for self-calibration. Furthermore, the poses of the projectors with respect to the objects center are initialized based on a uniformly-distributed subset of object points and optimized in a second step. The estimated poses are used for scale recovery. Finally, both reconstructions are combined using the high resolution of the photometric solution and the precise shape of the geometric solution.

Also worth mentioning are self-calibration techniques for laser line scanning. Winkelbach et al. proposed a self-calibration method for a hand-held laser line projector by placing the object in front of a corner with two known planes [9]. The system became popular as the Davidscanner and was acquired by Hewlett-Packard.

Furukawa and Kawasaki presented a method for self-calibration of a hand-held laser line projector and a fixed camera without any constraints to the geometry of the scene [10]. However, a laser line projector with a known metric configuration such as known angels between the planes is required in order to perform a 3D reconstruction. The latter approach is utilized and modified for the present system in this work and introduced later.

Methodology

To demonstrate the general functionality, a hardware setup as shown in Fig. is used. The laser projector LP-HFD2 [11] of the company Z-Laser and the IDS UI-5270CP Rev. 2 [12] are mounted as an active stereo system with a distance between 2 m and 3 m to the scene. The camera has a 1/1.8" CMOS sensor with 3.17 MP at a resolution of 2056x1542 with global shutter. The maximum shutter speed is 36 fps at full resolution. A Fujinon lens with a fixed focal length of 16 mm at an aperture of f5.6 is used. The projector uses a 7 mW fiber coupled green laser source and is designed for 2D and 3D projection as well as CAD drawings into a scene. It has an operating distance of 0.5 to 7 m and an accuracy of 0.25 mm/m related to the projection distance. The maximum projection frequency, depending on the projected scene, is given as 50 Hz. Several projectors can be used as a multi-projection system. Both devices are connected to a workstation (Intel Core i7-7820X, 64 GB RAM) via Gigabit Ethernet. To ensure time synchronization, the laser projectors trigger output is directly connected to the camera's trigger input.

The method applied for calibration and reconstruction is based on the self-calibration approach introduced by Furukawa and Kawasaki [10]. Although the original paper was written for a system consisting of a fixed camera and a hand-held cross-line laser projector which is moved through the scene, the methodology here is adapted for a system where both camera and projector are fixed. The movement is simulated by projecting randomized laser crosses all over the scene. To solve the problem of recovering scale from the scene, two possibilities were feasible. One

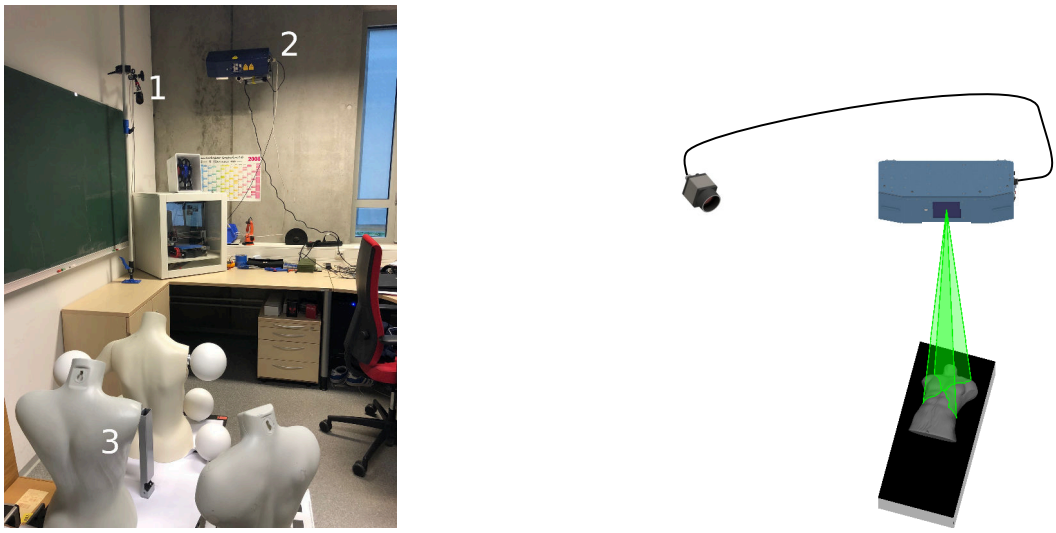

Figure 2. *Experimental Setup, Left: 1: Industrial Camera, 2: Industrial Laserprojector, 3: Calibration Scene, Right: Schematic Setup with Laser Cross and Trigger Connection.*

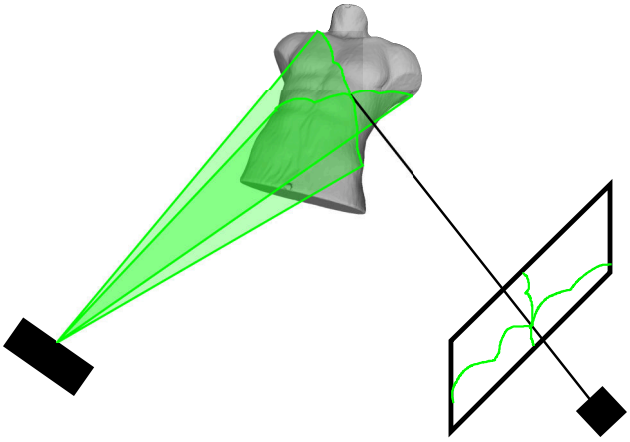

Figure 3. *Cross line laser projection as seen from a camera. Note, that both lines are projected consecutively and then merged to a cross during processing.*

Figure 4. *Sample scene with six calibration spheres. Phantoms are added for texture and variation.*

could use additional hardware such as a laser pointer attached and calibrated to the projector to get scaling from the relation between the two devices as used by Furukawa and Kawasaki in [7], or geometric figures of known shape could be introduced into the scene, hence providing additional constraints to recover scale. For this work, the latter approach was used. Calibration spheres for laser applications have an optimal reflective surface, are inexpensive and flexible with respect to positioning within the scene while also providing the necessary depth in structure needed for the approach. Also, no additional hardware is needed, which is a key point of this work, and no separate calibration step between the projector and the additional hardware is needed. Since the LP-HFD2 projector uses a class 2M green laser diode, both cross segments are projected consecutively to simplify the line extraction and further algorithmic processing. Once the grid of projected lines is reconstructed and thus, the system calibrated up to scale, reference spheres within the scene are detected and used for scaling. All plane equations determined this way can later be used to scan any arbitrary scene. A cross laser line projection as seen from a camera on a phantom is shown in Fig. 3.

This work demonstrates our approach only for a pair of camera and projector, it is however easily extendable to a multi-projector-camera system, as present in typical treatment rooms. The Winston-Lutz test calibrates the position of the isocenter of the linac and makes sure, that all cross-lasers are intersecting in the isocenter, hence making it visible. Since the position of the isocenter with respect to the linac and the laser projectors is known afterwards, the calibration performed in this work can be easily transformed into the linacs coordinate system. When calibrating more than one projector, using the linacs coordinate system as a reference provides the necessary information needed to combine everything into a multi-projector-camera system.

Before introducing the necessary theoretical background and describing our experiments, we are giving an overview of how, step by step, our experimental setup is used.

Camera Model

To model the physical properties of the camera, the well-known pinhole model is utilized. This gives a transformation from three-dimensional world coordinates (X,Y,Z) into two-dimensional image coordinates (u,v), following

$$(X,Y,Z)^T \mapsto \left(\frac{f_u X}{Z}+u_0, \frac{f_v Y}{Z}+v_0\right)^T = (u,v)^T, \tag{1}$$

where f_u and f_v are the cameras focal length parameters in the u and v direction and u_0 and v_0 as the cameras principle point. Additionally, radial (k_i) and tangential (p_i) distortion parameters are introduced as

$$\hat{\mathbf{x}} = \mathbf{x}(1+k_1r^2+k_2r^4+k_3r^6)+\begin{bmatrix}2p_1xy+p_2(r^2+2x^2)\\ p_1(r^2+2y^2)+2p_2xy\end{bmatrix} \tag{2}$$

with $r=\sqrt{x^2+y^2}$ as the distance of the point from the principal point and $\mathbf{x}$ as the distorted normalized point coordinates, to account for design imperfections and physical limitations. The camera is calibrated using Zhangs method [13].

Line Extraction

As previously described, the laser lines are projected in an arbitrary manner across the scene, with two consecutive lines forming one cross. Since the system is still uncalibrated and neither position nor orientation of the lines are known, simple line detectors such as peak detection along horizontal or vertical image direction are not applicable. To this end, Stegers line algorithm is used [14]. It must be noted, that the illumination during image capture needs to be as constant as possible, which is feasible to assume in a typical treatment room. To make the projected lines in the images more visible, an initially taken background image is subtracted from all acquired images.

The general idea of this algorithm is to detect lines, which have a characteristic 1D profile perpendicular to their direction, such as parabolic or Gaussian. Such profiles typically vanish for their first directional derivative and have a large absolute value for their second directional derivative. The direction of the line locally for each point is then computed by convolving the image I with the discrete two-dimensional Gaussian partial derivative kernels. The eigenvectors and values of the Hessian matrix

$$H(x,y)=\begin{pmatrix}\frac{\partial^2 g_\sigma(x,y)}{\partial x^2} & \frac{\partial^2 g_\sigma(x,y)}{\partial x\partial y}\\ \frac{\partial^2 g_\sigma(x,y)}{\partial x\partial y} & \frac{\partial^2 g_\sigma(x,y)}{\partial y^2}\end{pmatrix} * I(x,y) = \begin{pmatrix}r_{xx} & r_{xy}\\ r_{xy} & r_{yy}\end{pmatrix} \tag{3}$$

are calculated, where r_{xx}, r_{xy} and r_{yy} are the second partial derivatives. The direction perpendicular to the line $n=(n_x,n_y)$ with $\|(n_x,n_y)\|_2 = 1$ is then the eigenvector corresponding to the biggest eigenvalue. To determine whether the first directional derivative along (n_x,n_y) vanishes, a quadratic polynomial is used. The location of this point P along n is determined using a Taylor expansion, where the maximum or minimum is given by

$$(p_x,p_y) = (tn_x,tn_y) \tag{4}$$

with

$$t = -\frac{r_x n_x + r_y n_y}{r_b x n_x^2 + 2r_{xy}n_x n_y + r_{yy}n_y^2} \tag{5}$$

where r_x and r_y are the first partial derivatives. In order to be a point on the line and lie within the boundaries of a pixel, $(p_x,p_y) \in [-0.5,0.5]\times[-0.5,0.5]$ must hold. The point is determined with sub-pixel accuracy and is only valid, if the value of the maximum absolute eigenvalue is above a threshold, characterizing salient lines.

All identified points are then linked to line segments using the individual line directions and two thresholds for the second derivative for smoothing the final result and removing outliers. The directed search is described in more detail in [14] and will not be further described here.

Plane Parameter Estimation

This section introduces the plane parameter estimation of the uncalibrated system by Furukawa and Kawasaki [7]. Other than in the original paper, both the camera and the laser projector are fixed. Starting with a single laser line, it's plane (or curve) π can be represented as

$$\pi : aX + bY + cZ + 1 = 0. \tag{6}$$

Using the projective pinhole camera model from (1) and rearranging the terms yields

$$\pi : a\left(\frac{x-u_0}{f_u}\right) + b\left(\frac{y-v_0}{f_y}\right) + c = -\frac{1}{Z}. \tag{7}$$

Therefore, with known plane parameters $(a,b,c)^T$, the camera intrinsics matrix and the 2D coordinate $(x,y)^T$, the corresponding 3D point can easily be calculated using ray intersection as

$$\begin{aligned} Z &= -\frac{1}{a\left(\frac{x-u_0}{f_u}\right)+b\left(\frac{y-v_0}{f_y}\right)+c}\\ X &= Z\frac{x-u_0}{f_u}\\ Y &= Z\frac{y-v_0}{f_v}, \end{aligned} \tag{8}$$

with the Z coordinate negative, since the Z-axis is assumed to be directed backwards from the camera.

For two curves π_i and π_j with the intersection point $(x_{ij},y_{ij})^T$, combining the respective (7) as

$$\begin{aligned} -\frac{1}{Z} &= a_i\left(\frac{x_{ij}-u_0}{f_u}\right)+b_i\left(\frac{y_{ij}-v_0}{f_y}\right)+c_i =\\ -\frac{1}{Z} &= a_j\left(\frac{x_{ij}-u_0}{f_u}\right)+b_j\left(\frac{y_{ij}-v_0}{f_y}\right)+c_j \end{aligned} \tag{9}$$

yields

$$(a_i-a_j)\left(\frac{x_{ij}-u_0}{f_u}\right)+(b_i-b_j)\left(\frac{y_{ij}-v_0}{f_y}\right)+(c_i-c_j)=0. \tag{10}$$

(10) is homogeneous and contains the differences of the plane parameters. Therefore, both set of plane parameters have the same

indeterminacies, a scalar s and an offset vector $\boldsymbol{o}$. Two equations can be found as

$$\begin{aligned} \boldsymbol{a_i} &= (a_i, b_i, c_i)^T = s(a'_i, b'_i, c'_i)^T + \boldsymbol{o} \quad \text{and} \\ \boldsymbol{a_j} &= (a_j, b_j, c_j)^T = s(a'_j, b'_j, c'_j)^T + \boldsymbol{o}, \end{aligned} \tag{11}$$

where $\boldsymbol{a'_i} = (a'_i, b'_i, c'_i)^T$ and $\boldsymbol{a'_j} = (a'_j, b'_j, c'_j)^T$ are solutions for $\boldsymbol{a_i}$ and $\boldsymbol{a_j}$ up to scale. Given N curves with M intersections, all equations such as (11) are combined in a homogeneous linear system as

$$\mathbf{L}\boldsymbol{p} = 0 \tag{12}$$

with the $3N$-dimensional vector

$$\boldsymbol{p} = s\boldsymbol{A} + \boldsymbol{O}, \tag{13}$$

where $\boldsymbol{A} = (\boldsymbol{a'_1}, \ldots, \boldsymbol{a'_N})$ and $O = (\boldsymbol{o}, \ldots, \boldsymbol{o})$ as in (11), and the $M \times 3N$ matrix $\mathbf{L}$ containing $(\pm x_{ij} - u_0/f_u)$, $(\pm y_{ij} - v_0/f_v)$ and (± 1) at the appropriate positions to form the corresponding homogeneous linear equations as (10). The solution for $\boldsymbol{p_i}$ $(0 \leq i \leq N)$ is found as

$$\boldsymbol{p_i} = s(a'_i, b'_i, c'_i) + \boldsymbol{o} \tag{14}$$

having 4-DOF indeterminacy with a scalar s and an offset vector $\boldsymbol{o}$, if the system is solvable and has no degenerate conditions ($\geq$ 4-DOF indeterminacy, i.e. if all intersections for one curve a collinear) and is called *projective solution*. A trivial solution for $\boldsymbol{p}$ is obviously the zero vector, so the system (12) is solved using Singular Value Decomposition (SVD) under the constraint $\|\boldsymbol{p}\| = 1$. It is worth mentioning, that the 4-DOF indeterminacy of this general solution can be described as a 4 parameter homography, that transforms the 3D points and plane parameters.

To estimate the plane parameters up to scale, a metric constraint needs to be obtained. Since always pairs of two consecutive laser planes are projected perpendicular to each other, orthogonality is present. Given a set of pairs of orthogonal planes, an error function for each pair i, j from the set of all $C_v = (ij)|(\pi_i \perp \pi_j)$ is defined as

$$\begin{aligned} E(\boldsymbol{o}) &= \sum_{(ij)\in C_v} \cos^2 \theta_{ij}(\boldsymbol{o}) \\ &= \sum_{(ij)\in C_v} N(a_i, b_i, c_i, \boldsymbol{o})^T N(a_j, b_j, c_j, \boldsymbol{o})^2, \end{aligned} \tag{15}$$

where θ_{ij} is the angle between the two planes and $N(\cdot)$ is the normal of the plane computed using plane parameters and the offset vector. A vector $\hat{\boldsymbol{o}}$ is found using non-linear optimization, which minimizes the error function as

$$\hat{\boldsymbol{o}} = \underset{\boldsymbol{o}}{argmin} E(\boldsymbol{o}). \tag{16}$$

This solution is called *metric solution* and calibrates the system up to scale.

Scaling the Point Cloud

To recover scale, a metric constraint needs to be obtained. This is done by introducing test specimen with a known size into the scene. For this purpose, laser scanning reference spheres with a diameter of 145 mm are used. They are precisely manufactured and coated with an optimal reflective paint for laser scanning applications. Between three and five of those reference spheres are randomly placed in the scene. After the reconstruction of the point cloud up to scale, the following steps are performed. A k-d tree representation of the input data is created to speed up the processing. Downsampling is performed according to the point cloud dimensions. Next, an Euclidean Cluster Extraction is used for segmentation. Finally, for all found segments, a sphere model is fitted with RANSAC and the radii are determined.

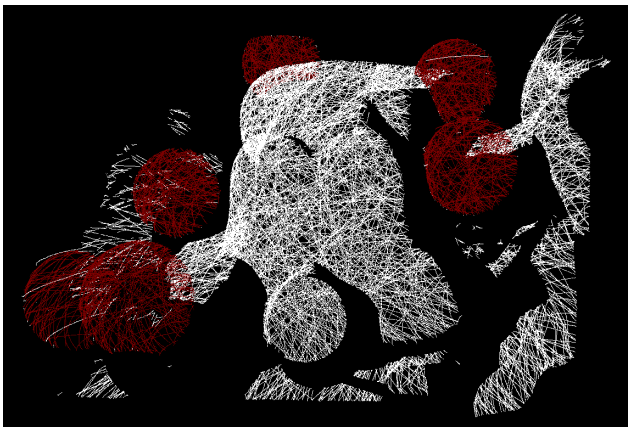

Figure 5. *Segmentation of Point Cloud, detected spheres are colored in red.*

By comparing the determined radii and knowing the number of spheres in the scene, non-sphere segments are marked as outliers. A segment is considered as an actual sphere, if the RMS of the fitting is less than 1 and the difference to all other sphere radii is not bigger than 5 %. Finally, all identified sphere radii are averaged and the relation to the real sphere radius of 72.5 mm gives the final scale factor. To be able to use for all steps the same empirically determined parameters despite the varying scale, downsampling is performed according to the extent of the point cloud. Fig. 5 shows a sample point cloud with all detected segments. For scaling and segmentation, the PCL framework [15] is used.

Experiments

Sphere Detection and Scaling

To determine the performance of the self calibration with scaling, two sample scene similar to Fig. 4 are set up. Three and six calibration spheres ($r = 72.5$ mm) are randomly placed inside the scenes. For each scene, ten consecutive scans with 300 lines (=150 randomized crosses) and 600 lines (=300 randomized cross) for the scene with three and six spheres, respectively, are projected and automatic scaling is performed. Camera shutter speed is set to 39.74 ms. Finally, the scaled point cloud is again used to fit virtual spheres in the reconstructed data to obtain the scaled radii of the calibration spheres. The overall procedure took around 40 s for 150 crosses and 85 s for 300 lines for projection, line extraction, self-calibration and scaling.

Fig. 6 shows box plots of the determined radii per scan. From both plots it is visible, that most scans result in a mean sphere radius which is centered around the reference radius ± 0.5 mm. The

range, where the radii vary within, goes from around 1.5 mm in scan number 6 up to 5.5 mm in scan number 3 (Fig. 6, top). Naturally, a small difference from the biggest to the smallest determined radius as well as a centering around the reference radius is desired. From both plots it is also clear, that there is no clear visible scheme for the outcome of the determined radii which could be caused by systematic errors. Although box plots themselves do not give any information which could lead to the exclusion of systematical errors, since all points are plotted, and no outliers are visible. Also, the minima and maxima between consecutive scans do not lie close enough to each other, that a general shift for one of the spheres can be assumed. Fig. 7 shows box plots of the determined sphere radii from 10 scans, separate for each sphere. Again no clear pattern of shift is visible. The plots are done to test the scaling for repeatability. Especially for spheres number 1, 3 and 4 (Fig. 7, top) the deviation of around 3 mm lies within the range which is for our system in its current state acceptable, as it is a deviation of 4.13 % around the ideal radius of 72.5 mms. The plots show no significant degradation of the results when using six instead of three spheres. For robustness in practical applications however a higher number of spheres increases robustness in case of occlusions.

Considering the occurring offsets in all box plots following no pattern, two reasons come to mind. First, Fig. 8 shows all projected and extracted lines with detected intersection marked green. It is clearly visible, that a high amount of camera pixels is not illuminated due to a limited field of view overlap between camera and projector. This is a result of the fixed base line compared to a varying base lines from the original proposed self-calibration paper. As a result, the reconstructed segments of the reference spheres, which are used for scaling, represent only less than half of a real sphere. Second, the line distribution on the sphere segment itself is not uniform (especially visible in Fig. 8, top left corner). Together, this introduces random errors into the RANSAC based sphere fitting, which are prone to negatively impact the result.

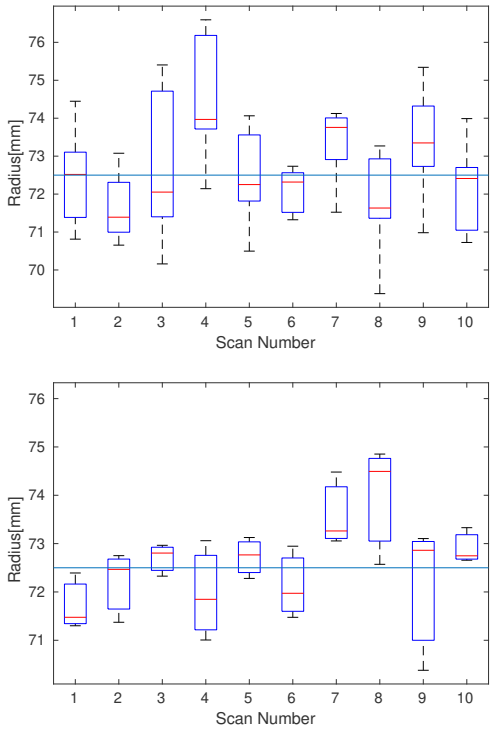

Figure 6. *Determined sphere radii per scan for 6 (top) and 3 (bottom) spheres within the scene. The vertical line marks the actual sphere radius.*

Shape Acquisition and Verification

The previously determined plane equations are further on used for shape acquisition. For both calibrated systems with 150 and 300 crosses as introduced in section *Sphere Detection and Scaling*, two phantoms are scanned and their surface is reconstructed. To evaluate the reconstruction, reference scans of the two test phantoms were acquired using an Ensenso N35 3D Vision Camera [16] with sub millimeter accuracy. Our reconstructions are registered with the reference using CloudCompare's integrated tools. Additionally, the point to point distance is calculated. A typical output point cloud of the Ensenso's reference scan has around 1 million points compared to between 100000 and 300000 points from the shape acquisition, depending on the number of crosses used. Fig. 9 shows both acquired point clouds, colored with the point to point error, left and right respective to Fig. 6 and Fig. 7. The color map ranges from 0 to 3 mm. As can be seen, both point clouds contain areas, where the point to point error is exclusively 3 mm. In those areas, the ground truth scans do not contain any reference data, hence any errors of 3 mm or greater are not considered in the following.

From a qualitative analysis, the error distribution within the point cloud is regular, which makes the reconstruction useful for

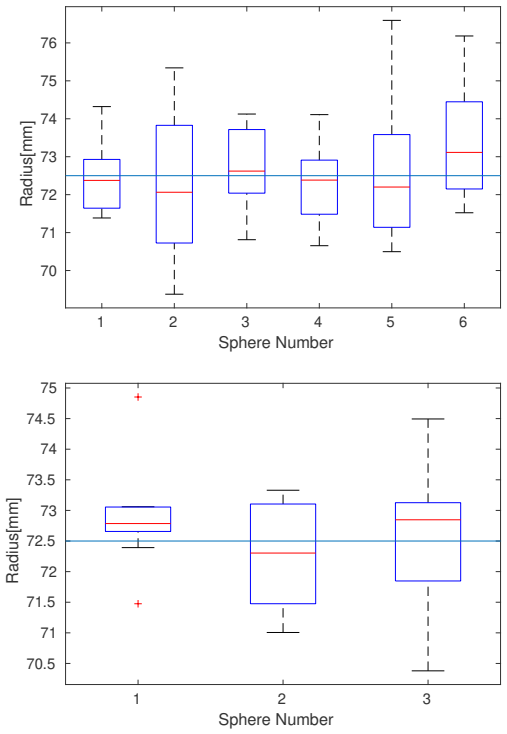

Figure 7. *Determined sphere radii per sphere for 6 (top) and 3 (bottom) spheres within the scene. The vertical line marks the actual sphere radius.*

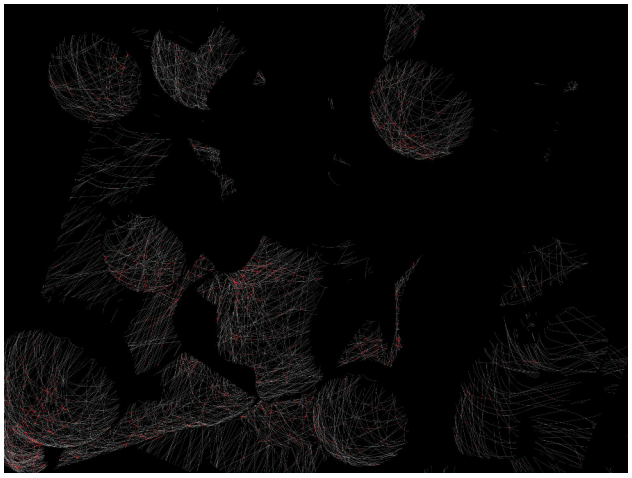

Figure 8. *Image of all extracted lines from a scan. Intersections used for reconstruction are marked red.*

position verification and correction, since the shifting vector between current and actual position is not affected by inconsistencies within the point cloud. Fig. 10 shows the point to point error for the left image from Fig. 9 as a histogram. Again, all errors greater than 3 mm are summarized in one box and not considered. It is visible, that most of the errors lies below 1.5 mm. The heavily tailed distribution of the errors gives the conclusion, that multiple outliers are present. These could either be, as mentioned before, still point where no ground truth data is available, or be the result of errors introduced along the processing pipeline starting with the line extraction. Naturally, in such specific and demanding areas it is necessary to keep errors as small as possible, on one hand by improving the integrity of the system in general, but on the other hand also by raising the thresholds for deciding if a point is considered to be valid or an outlier.

From this analysis it is feasible to state that, despite the problems with scaling described in the previous section, the reconstruction of the surface performs well for a system which is still work in progress. However, we would like to emphasize again that we are working on improving the calibration and reconstruction to achieve the required accuracies in medical processes.

Conclusion

Summarizing the paper, a few key aspects are noteable:

1. The application of the self-calibration methodology introduced by Furukawa and Kawasaki for a static laser projector was shown.
2. Extending the original approach, a reconstruction with scale was done using reference spheres within the calibration scene.
3. The hardware setup used is similar to the already existing hardware in medical treatment rooms and the methodology proposed is therefore applicable for shape acquisition and position verification of patients without the need of an additional external system.
4. By combining and extending different approaches we showed that a different approach with readily available hardware for patient positioning is feasible. The demonstration of the full potential of the system such as the implementation of respiratory gating is still work in progress.

But needless to say, a lot of work remains to be done. Improving the robustness of the self-calibration with respect to the repeatability is a key aspect for a productive use. Since the projector is fixed, only small sections of the reference spheres are actually visible in both fields of view and shadows are present in many areas of the scene. A possible solution to this problem is to use the initially known plane parameters of the projected laser lines in the projector coordinate system and a initial estimation of the extrinsics via shape projections other than lines as a starting point for the self-calibration and further refinement. Also, extending the approach to a multi-camera-projector system could on the other hand solve the weak coverage problem when illuminating the scene from different angles. Finally, the calibration scene used in this paper is hardly practical. Thus a calibration pattern needs to be designed, which provides the following key features. It must have enough metric constraints such as spheres or cubes of known sizes to allow for robust scaling. Additionally, it must be applicable for various physical hardware configurations in different treatment rooms, while minimizing shadows.

Acknowledgments

Gefördert durch: Bundesministerium für Wirtschaft und Energie aufgrund eines Beschlusses des Deutschen Bundestags

References

[1] David. P. Gierga et al, Analysis of setup uncertainties for extremity sarcoma patients using surface imaging. Practical radiation oncology 4.4: 261, (2014).

[2] Christoph Bert et al., A phantom evaluation of a stereovision surface imaging system for radiotherapy patient setup., Medical physics 32.9: 2753, (2005).

[3] Guang Li et al., Characterization of opticalsurfaceimagingbased spirometry for respiratory surrogating in radiotherapy., Medical physics 43.3: 1348, (2016).

[4] visionrt, AlignRT, http://www.visionrt.com/product/alignrt/.

[5] C-RAD, Catalyst HD, https://c-rad.se/product/catalyst-hd/.

[6] Anzai Medical Co. LTD, Respiratory Gating System AZ-773VI, http://www.anzai-med.co.jp/en/product/item/az733vi/index.html.

[7] Ryo Furukawa and Hiroshi Kawasaki, Uncalibrated multiple image stereo system with arbitrarily movable camera and projector for wide range scanning, IEEE Proc. 3-D Digital Imaging and Modeling, 2005, pg. 302–309. (2005).

[8] Daniel G. Aliaga and Yi Xu, A self-calibrating method for photogeometric acquisition of 3D objects, IEEE Transactions on Pattern Analysis and Machine Intelligence, 32, 4 (2010).

[9] Simon Winkelbach and Sven Molkenstruck and Friedrich M Wahl, Low-cost laser range scanner and fast surface registration approach, Lecture Notes in Computer Science, 4174, 718 (2006).

[10] Ryo Furukawa and Hiroshi Kawasaki,Self-calibration of multiple laser planes for 3D scene reconstruction, IEEE 3D Data Processing, Visualization and Transmission, 200-207, (2006).

[11] Z-Laser Optoelektronik GmbH, LP-HFD2, https://z-laser.com/en/product/laser-projector/lp-hfd2/.

[12] IDS, UI-5270CP Rev. 2, https://en.ids-imaging.com/store/ui-5270cp-rev-2.html.

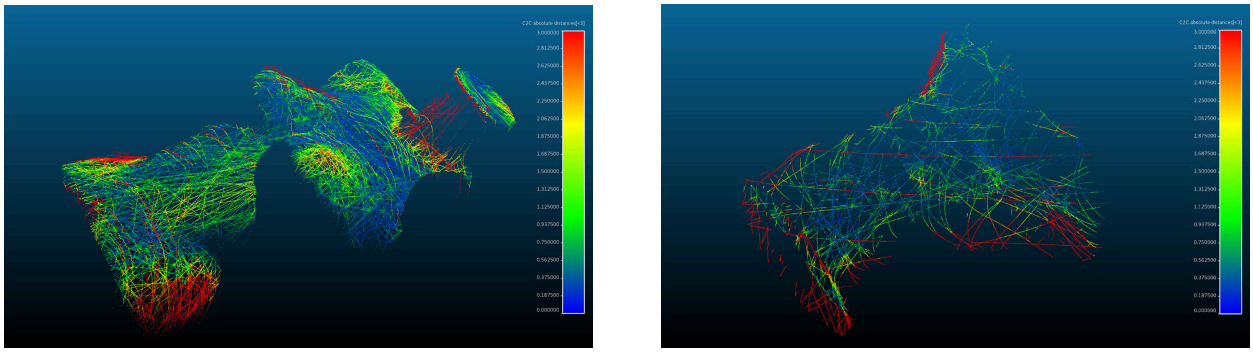

Figure 9. *Scans colored with point to point distance [mm] to reference scan. The coloring scale varies from 0 to 3 mm. Left: Calibration with 6 spheres and scan with 300 crosses (respective to Fig. 6 top, scan 6), Right: Calibration with 3 spheres and scan with 150 crosses (respective to Fig. 6 bottom, scan 3).*

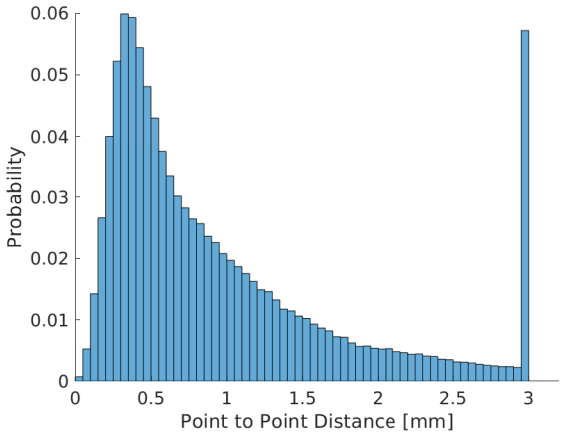

Figure 10. *Histogram of the point to point error after registration, respective to Fig. 6 top, scan 6.*

[13] Zhengyou Zhang, A Flexible New Technique for Camera Calibration, IEEE Trans. Pattern Anal. Mach. Intell, 1330-1334. (2000).

[14] Carsten Steger, An unbiased detector of curvilinear structures., IEEE Transactions on Pattern Analysis and Machine Intelligence, 113125 (1998).

[15] Radu Bogdan Rusu and Steve Cousins, 3D is here: Point Cloud Library (PCL), IEEE International Conference on Robotics and Automation (ICRA), (2011).

[16] 3d Camera Ensenso N35, https://en.ids-imaging.com/ensenso-n35.html.

Author Biography

Sven Jörissen received his B.Sc. in Aerospace Information Technology from University of Würzburg (2015) and his double M.Sc. degree in Space Science and Technology from University of Würzburg and Lulea University of Technology (2017). He is currently pursuing his PhD studies as a research associate at University of Würzburg within the field of industrial and medical laser scanning. His research interests include robotics, sensor data processing and 3D point clouds.

Michael Bleier received his diploma in Mechatronics Engineering from the Friedrich-Alexander University Erlangen-Nuremberg (2013). Since 2014 he is working as a research associate at Zentrum für Telematik in Würzburg. At the research institute he is working in the context of sensor data processing and underwater laser scanning. His research interests include 3D computer vision, robotics and underwater mapping.

Andreas Nüchter is professor of computer science (telematics) at University of Würzburg. He holds a doctorate degree (Dr. rer. nat) from University of Bonn. His main research interests include reliable robot control, 3D environment mapping, 3D vision, and laser scanning technologies, resulting in fast 3D scan matching algorithms that enable robots to perceive and map their environment in 3D representing the pose with 6 degrees of freedom.