

# Accuracy of object tracking based on time-multiplexed structured light

---

B. Wagner<sup>1,2</sup>, P. Stüber<sup>1,2</sup>, T. Wissel<sup>1,2</sup>, R. Bruder<sup>1</sup>, A. Schweikard<sup>1</sup>, F. Ernst<sup>1</sup>

<sup>1</sup> University of Lübeck, Institute for Robotics and Cognitive Systems, Lübeck, Germany

<sup>2</sup> University of Lübeck, Graduate School for Computing in Medicine and Life Sciences, Lübeck, Germany

Contact: wagner@rob.uni-luebeck.de

## Abstract:

*Our research group is currently developing a new optical head tracking system which utilises infrared laser light to measure features of the soft tissue on the patient's forehead. These features are intended to offer highly accurate registration with respect to the rigid skull structure by means of compensating for the soft tissue. In this context, the system also has to be able to quickly generate accurate reconstructions of the skin surface. For this purpose, we have developed a laser scanning device which uses time-multiplexed structured light to triangulate surface points. This paper shows, that time-multiplexed structured light can be used to generate highly accurate reconstructions of surfaces (RMS error 0.17 mm, Kinect: RMS error 0.89 mm). Moreover, we used our laser scanner for tracking of a rigid object to determine how this process is influenced by the remaining triangulation errors. It turned out that our scanning device can be used for high-accuracy tracking of objects (RMS errors of 0.33 mm and 0.12 degrees).*

*Keywords: optical head tracking, time-multiplexed structured light, triangulation*

## 1 Introduction

Image-guided intracranial radiosurgery relies on highly accurate tracking of the patient's head. This can be achieved by using the rigid skull structure for registration between a reference imaging dataset and imaging datasets that are acquired during treatment. Accuray Inc. (Sunnyvale, CA, USA) developed the 6D Skull Tracking [1] for their CyberKnife system. This method uses stereoscopic X-ray images in which the rigid bone structure of the head is segmented and registered to a planning CT. Experiments with an anthropomorphic skull phantom revealed that this method achieves sub-millimeter tracking accuracy. The On-board *Imager (OBI)* by Varian Medical Systems Inc. (Palo Alto, CA, USA) uses cone-beam CT for target localisation regarding bone structures. According to [2], experiments with an anthropomorphic head phantom showed that the OBI system also achieves sub-millimeter tracking accuracy. The high accuracy of X-ray based tracking methods is achieved at the cost of tracking speed. This is due to the fact that X-ray imaging has to be carried out at low frame rates to guarantee acceptable additional dose levels.

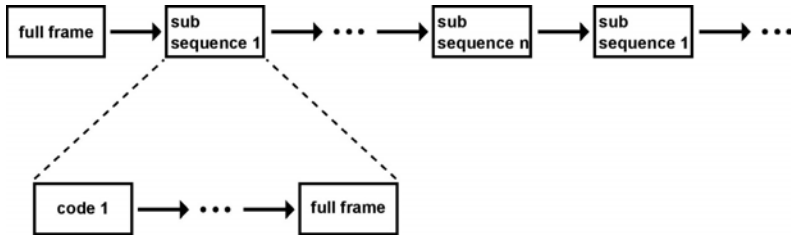
By contrast, optical head tracking offers much faster scan rates. Here, visible or infrared light is actively or passively used to reconstruct the surface of the skin. However, this method is highly inaccurate since the registration process is influenced by deformable structures which are given by the skin and the muscles. In order to overcome this problem, we are currently developing a system which utilises infrared laser light to measure features of the soft tissue on the patient's forehead [3]. These features are intended to offer highly accurate registration with respect to the rigid skull structure by means of compensating for the soft tissue. In this context, the system also has to be able to quickly generate accurate reconstructions of the skin surface. For this purpose, we have developed a laser scanning device which uses time-multiplexed structured light to triangulate surface points [4]. In this paper, the triangulation accuracy of our laser scanning device is analysed. Since Microsoft's Kinect represents an alternative for fast surface reconstruction, we also compare our results to the triangulation accuracy of the Kinect device. Furthermore, we used our developed laser scanner for tracking of a rigid object to determine how this process is influenced by the remaining triangulation errors. For this measurement, the scanning device was mounted to the end-effector of a robot to be able to calculate a ground truth for the tracking. Finally, the tracking accuracy was computed for a set of 200 tracking results.

## 2 Material and Methods

The laser scanning device generates a red laser point which is redirected by two moveable mirrors to create specific laser point patterns on a target object. An IDS UI-3240CP-NIR-GL camera captures all laser points pattern wise. The laser scanning device was configured to project a grid of 56 x 72 equidistant laser points at a frame rate of 10 Hz. Template matching based on normalized cross-correlation is used for highly accurate detection of the centres of the imaged laser

points. In order to be able to reconstruct the surface of a target object, every laser ray was calibrated with respect to the camera. Therefore the laser grid was projected onto a planar calibration body which was placed at different distances with respect to the laser scanning device. By means of the calibration body, the homography matrix [5] was calculated to map all imaged laser points to spatial points relative to the calibration body. This was done for every projected grid and afterwards all spatial points were used to fit a 3D line for every laser ray. The final calibration step was given by defining the pose of every laser ray with respect to the camera.

The reconstruction of the surface of a target object requires knowledge about the mapping between the laser points in an image and the matching calibrated laser rays. This is also known as the correspondence problem. Due to the field of view and the shadowing of the camera, it can't be guaranteed that all laser points of the projected pattern are captured by the camera. Hence, the correspondence problem can't be solved directly in the image. For this reason, time-multiplexed structured light [6] has been utilised in this work. The basic idea of this method is that within  $n$  frames, every laser ray projects a unique code for its identification. In this work, an extended version of the binary code is used for this task. Since the target object will move during head tracking, a high number of consecutive zeros has to be avoided in the code sequences to ensure the tracking of the laser points in an image sequence. For this reason, the binary code was extended by so-called full frames. When the camera captures a full frame it is ensured that all laser points of the pattern are projected. As shown in Fig. 1, the process starts with a full frame to get the initial location of the laser points that are visible in the image. Subsequently, every image sequence is arranged as sub sequences. Every sub sequence has a configurable length of images and always ends with a full frame. In this context, a length of two images means that a sub sequence includes one code frame which is followed by a full frame. Furthermore, the code frames start with the least significant bit and end with the most significant bit. After the first  $n$  sub sequences, all tracked laser points have been identified and consequently a reconstruction can be carried out for the final full frame. Since the identified points are tracked in the following images, a reconstruction can also be carried out for every following full frame. Moreover, points that were not visible before are considered by the identification process as soon as they are visible at the beginning of a sequence.

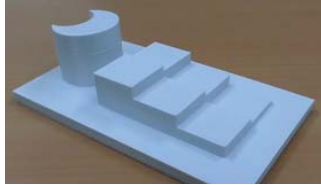


**Figure 1:** Extension of the code sequence by full frames

The reconstruction of the surface of a target object is carried out by triangulation of spatial laser points with respect to the camera. The used method for the triangulation of a spatial laser point is referred to as the Linear-Eigen method and depends on the respective laser point in the image and the corresponding calibrated laser ray. For triangulation in general, it applies that incorrectly identified laser points in the image lead to outlying spatial points. To avoid this, the identification of a laser point in the image is verified by means of the corresponding calibrated laser ray. Therefore the calibrated laser ray is projected in the image plane. The resulting line in the image is called epipolar line. Afterwards, the identification of the laser point in the image is verified by calculating the perpendicular distance of the point to the epipolar line. The point is considered to be identified correctly if the perpendicular distance does not exceed a threshold of 1 pixel.

To analyse the triangulation accuracy of the developed laser scanning device, two experiments have been carried out. In the first experiment, a planar surface was reconstructed to calculate the deviation of the resulting points with respect to the plane. For this purpose, the resulting point cloud was processed with Principal Component Analysis (PCA) to calculate the three principal axes of the point cloud. Subsequently, the point cloud was transformed by using the rotation matrix that is given by the three computed principal axes. Afterwards, the mean-free equivalent of the transformed points was calculated. Consequently, the sought deviation of the triangulated points was given by the point data regarding only one axis. Since Microsoft's Kinect represents an alternative for fast surface reconstruction (reconstruction rate of 30 Hz), we also compare our results to the triangulation accuracy of the Kinect device.

In the second experiment, the overall accuracy of the triangulation was determined by considering the deviations in all axes. For this purpose, a plastic triangulation phantom was designed (see Fig. 2). Subsequently, a ground truth for the reconstruction of the phantom was created. This ground truth was given by a CT-based point cloud of the phantom’s surface (Siemens SOMATOM® Definition AS+, voxel size 0.359 x 0.359 x 0.6 mm<sup>3</sup>). Afterwards, the triangulation phantom was reconstructed by using our laser scanning device. In order to calculate the overall accuracy of the triangulation, the resulting point cloud was registered to the ground truth by using the Iterative Closest Point algorithm (ICP) [7]. Here, the *point-to-plane distance metric* was utilised since *different reconstructions of the same object contain only few or no corresponding points*. After registration, the overall triangulation accuracy was given by the remaining distances between points and planes. Again our results were compared to the overall triangulation accuracy of the Kinect device.



**Figure 2:** Triangulation phantom

To determine the influence of the remaining triangulation errors on tracking, we used our laser scanner for tracking of a rigid object (here given by the triangulation phantom). For this measurement, the scanning device was mounted to the end-effector of an Adept Viper s850 robot. This approach allows the calculation of a ground truth for the tracking. The tracking itself is realized by using the ICP algorithm for the registration of a 3D reference point cloud to a second 3D point cloud. Here, *the* point-to-plane distance metric was used again. The result of the registration is given by a rotation matrix  $R$  and a translation vector  $t$  which define the estimated pose of the object with respect to the camera. After the acquisition of the first reconstruction (reference point cloud) of the surface of the object, consecutive translational displacements of 0.2 mm were applied to the robot end-effector. The space which includes all applied end-effector displacements is described by a sphere (with radius  $r = 6$  mm) and the initial pose of the end-effector defines the centre of the sphere. After the completion of a displacement, the laser point pattern was projected onto the surface of the object and the camera captured a new image. Since the imaged laser points were already identified after the acquisition of the reference point cloud, a new reconstruction of the surface could be calculated for every new full frame. Concerning the described method for time-multiplexed structured light, a length of two images was utilized for the sub sequences. For tracking, the reference point cloud was registered to all new reconstructions. Finally, the ground truth of the tracking was given by the respective translational end-effector displacements. Based on this procedure, the tracking accuracy was computed for a set of 200 tracking results.

### 3 Results

Tab. 1 shows the triangulation accuracy which is given by the root mean square (RMS) of the deviation of the reconstructed points with respect to a planar surface. Tab. 2 shows the overall accuracy of the triangulation which is defined by the RMS of the remaining distances between points and planes after registration with the ground truth. Tab. 3 shows the accuracy of the tracking of a rigid object for a set of 200 tracking results.

Device	RMS point-to-plane distances [mm]
Our laserscanner	0.1119
Kinect	0.7580

**Table 1:** Triangulation accuracy for the reconstruction of a planar surface

Device	RMS point-to-plane distances [mm]
Our laserscanner	0.1649
Kinect	0.8905

**Table 2:** Overall accuracy of the triangulation

RMS translational error [mm]	0.3825
RMS rotational error [degree]	0.12

**Table 3:** Tracking accuracy

## 4 Discussion

The results in Tab. 1 and Tab. 2 show that our laser scanning device clearly outperforms the Kinect device as far as the accuracy of the triangulation is considered. One major reason for the high accuracy of the triangulation is given by the use of time multiplexed structured light. This method ensures that the number of incorrectly identified laser points in an image is very low. Furthermore, identified points are verified by means of their corresponding epipolar lines. The results in Tab. 3 show that the remaining triangulation errors only cause small inaccuracies concerning the tracking of a rigid object. Hence, the reconstructed point clouds can be used for highly accurate tracking of objects.

## 5 Conclusion

Currently, we are developing a new optical head tracking system which utilises infrared laser light to measure features of the soft tissue on the patient's head. These features are intended to offer highly accurate registration with respect to the rigid skull structure by means of compensating for the soft tissue. In this context, the system also has to be capable of a fast generation of accurate reconstructions of the skin surface. For this purpose, we developed a laser scanning device which uses time-multiplexed structured light to triangulate surface points. This paper shows, that time-multiplexed structured light can be used to generate highly accurate reconstructions of surfaces. Since Microsoft's Kinect represents an alternative for fast surface reconstruction, we also compare our results to the triangulation accuracy of the Kinect device. The results show that our laser scanning device outperforms the Kinect device by a factor of five. To determine the influence of the remaining triangulation errors on tracking, we used our developed laser scanner for tracking of a rigid object. It turned out that the remaining triangulation errors only cause small inaccuracies for the tracking. In future works, the presented tracking shall be improved by using a stochastic filter which offers a better compensation of noise over time.

## 6 Acknowledgements

This work was supported by Varian Medical Systems Inc. (Palo Alto, CA, USA). Furthermore, this work was supported by the Graduate School for Computing in Medicine and Life Sciences funded by Germany's Excellence Initiative [DFG GSC 235/1].

## 7 References

- [1] Fu, D.; Kuduvalli, G.; Mitrovic, V.; Main, W.; Thomson, L.: *Automated skull tracking for the CyberKnife image-guided radiosurgery system*. Proc. SPIE 5744, Medical Imaging 2005: Visualization, Image-Guided Procedures, and Display, April 2005
- [2] Kim, GY; Pawlicki, T.; Le, QT; Luxton, G.: *Linac-based on-board imaging feasibility and the dosimetric consequences of head roll in head-and-neck IMRT plans*. Medical Dosimetry, Volume 33, Issue 1, Spring 2008, pages 93-99
- [3] Ernst, F.; Bruder, R.; Wissel, T.; Stüber, P.; Wagner, B.; Schweikard, A.: *Real time contact-free and non-invasive tracking of the human skull - first light and initial validation*. Proceedings of SPIE Optical Engineering + Applications, SPIE Optics + Photonics, SPIE, August 2013
- [4] Wagner, B.; Stüber, P.; Wissel, T.; Bruder, R.; Schweikard, A.; Ernst, F.: *Time-multiplexed structured light for head tracking*. 44. Jahrestagung der Deutschen Gesellschaft für Medizinische Physik, DGMP, September 2013
- [5] Hartley, R.; Zisserman, A.: *Multiple View Geometry in Computer Vision*. 2nd edition. Cambridge University Press, 2003. – ISBN 0-521-54051-8
- [6] Salvi, J.; Fernandez, S.; Pribanic, T.; Llado, X.: *A state of the art in structured light patterns for surface profilometry*. Pattern Recognition, Volume 43, Issue 8, August 2010, pages 2666-2680
- [7] Besl, P.J.; McKay, H.D.: *A method for registration of 3-d shapes*. IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 14, Issue 2, 1992, pages 239–256