Alignment of multiple Time-of-Flight 3D Cameras for Reconstruction of walking Feet

[1,2,3] Michael Stürmer, [1]Claude Seiler, [1]Guido Becker and [2,3]Joachim Hornegger
[1]Lion Systems S.A., Luxembourg,
[2]Pattern Recognition Lab, Department of Computer Science, Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany
[3]Graduate School of Advanced Optical Technologies (SAOT), Erlangen, Germany

INTRODUCTION

Three dimensional information about the surface of feet is valuable for many applications. Orthopedics, sport scientists and shoe manufacturers are interested in comparing the shapes of feet and to study the deformation when the load on the foot changes. While for the time being there are only static acquisition devices commercially available, several research groups are investigating the possibilities to go to the next level: 3D acquisition of walking feet.

All technologies share in common, that they need several fast acquisition devices. Each of these devices records a subset of the point cloud that will be forming the final reconstructed foot. A crucial step, before reconstruction of the foot and before anatomical information can be extracted by the physician, is the accurate alignment of the separate datasets.

We propose a method to align 3D datasets from an arbitrary number of cameras. A cubic reference object with known geometry is acquired from several viewpoints. For alignment angular and distance relations between the extracted faces from the recordings are integrated into a nonlinear optimization functional. This method is robust against noise within the data and can be fully automated. Mean rotational and translational errors of the alignment are 0.8° and 0.22mm.

Our device is using Time-of-Flight (ToF) cameras for acquisition of the 3D data. These special cameras provide with a lock-in pixel structure a markerless and dense 3D acquisition of objects with up to 50 frames per second [1]. Thus they lend themselves to be used as basis for dynamic scanning devices which generate 3D models with circumferential visibility.

METHODS

Registration (or extrinsic calibration) of cameras is a common task in computer vision and has been investigated in the past [2]. All methods are based on local feature extraction, i.e. only few pixel values contribute to one feature. The accuracy of the positions of the features in the images steer the quality of the reconstructed 3D data and the resulting registration. Even though ToF cameras provide up to 25k 3D points per image, their lateral resolution is low compared to ordinary 2D cameras. Accurate local feature detection is not always possible in this condition. Additionally, if at feature positions too much noise in the 3D data is present, robust 3D point correspondences between the separate datasets cannot be established. In our approach we compute features, each of which incorporate as many 3D points as possible. Thus the influence of local noise at specific feature positions is reduced.

The proposed setup involves three SR4000 ToF cameras which are seeing a cubic object from different directions (Figure 1). Every camera sees different parts of the object and

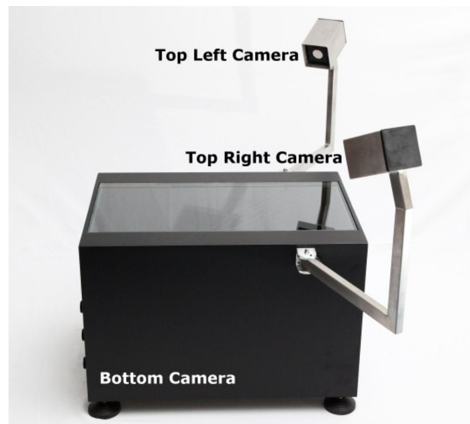

Figure 1: Proposed setup with multiple ToF cameras. The bottom camera is placed below the glass plate.

provides a separate dataset. From the known geometry and orientation of the block, for every dataset a defined number of planes P_i can be extracted from the 3D data (Figure 2). Each plane is defined by a point p_i (the mean of all 3D points contributing to P_i) and a normal vector n_i, where i denotes the number of the plane (1-8).

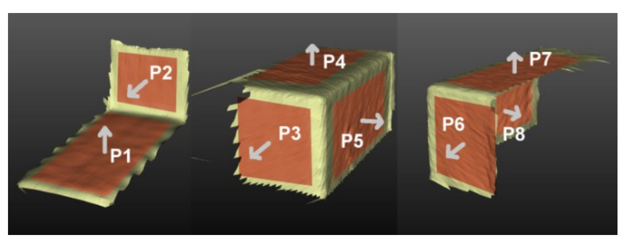

Figure 2: Planes extracted from the datasets of the cameras. Arrows show the normal vectors. From left to right: bottom camera, top left camera, top right camera.

For a given pair of two planes from one selected dataset it is known that the angle between the normals of these planes is 90°. Also the dimensions of the object are known. With this knowledge about angular and distance relations it is possible to define a set of constraints, the values of which need to be minimized to create a good registration T. T consists of three rigid transformations τ_{bottom}, τ_{left} and τ_{right}. Each transform is defined by three rotational and three translational parameters. One of these transformations is used as reference and thus need not to be calculated. We use τ_{bottom} as reference. Consequently a set of six parameters for τ_{left} and τ_{right} has to be found to define the final values of T. Before computation of the minimization constraints, the planes P_i must be transformed using T. A transformed plane P_i' with normal n_i' and point p_i' denotes the plane P_i transformed by τ_c, where c is the transform index (*bottom*, *left* or *right*) corresponding to the dataset from which P_i originates.

Two types of constraints are defined; angular constraints contribute to the rotational parameters of T while distance constraints define the translational parts of T. Angular constraints involve the normals n'_i and n'_j of two planes from different datasets. The remaining angular error e_a^{ij} is computed as

$$e_a^{ij} = |\arccos\left(\overrightarrow{n'_i} \cdot \overrightarrow{n'_j}\right) - \alpha_{ij}|,$$

where α_{ij} is the ground truth angle between the planes i and j. Distance constraints can only be derived from parallel planes of distinct datasets. For a pair of planes i and j the distance error e_d^{ij} computes as

$$e_d^{ij} = |\ \left[\ \left|n'_i \cdot \left(p'_j - p'_i\right)\right|\ \right] - d_{ij}\ |,$$

d_{ij} being the actual distance between the planes. For a set of M angular and N distance constraints the error e, that needs to be minimized to obtain the final values of T, results in

$$e = \sum_{k=1}^{M} e_a^k + \sum_{l=1}^{N} e_d^l.$$

The indices k, l denote all valid combinations of plane indices for angular and distance constraints. The final transformation set T is computed with a standard levenberg-marquard approach that solves the nonlinear least squares problem described by the above equation [3].

RESULTS AND DISCUSSION

After extraction of the planes $P_{1,2\ldots8}$ (see Fig. 2), the mean angular error between the planes from same datasets was 0.4°. This error results from imperfect intrinsic calibration of the cameras, and measurement inaccuracies. It influences the global minimum of the optimization functional, the smaller the error is, the smaller the final value of the function will be.

P_i were calculated from 500-4300 3D points, on average for each plane 1800 points where used. The noise of the plane data in normal direction was 0.40mm on average.

After minimization, the mean value of all e_a was 0.8°, while e_d showed values of 0.22mm on average. The result after applying T is shown in Figure 3. Using the proposed registration approach, we are able to align datasets of recorded feet and to do an accurate reconstruction of the 3D shape. Comparison of our dynamic scanning device with ground truth measurements from human feet have shown that the accuracy of the system is within the requirements for biomechanical applications (Figure 3) [4].

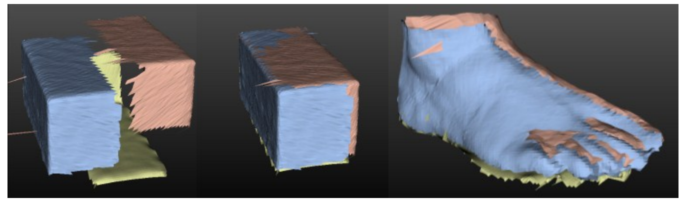

Figure 3: The reference dataset used for setting up the registration. Left: the unregistered datasets, Middle: the final registered object. On the right the aligned datasets from a foot are shown.

CONCLUSIONS

We have shown that we are able to align the data of several ToF cameras such that accurate reconstruction of objects of the size of feet is possible. The mean angular and translational errors after alignment are 0.8° and 0.22mm.

Future work will include automatic plane detection within the recorded datasets and enhancement of the method to work with several recordings. This will increase the number of constraints in the optimization and improve the robustness of the method. In addition the approach will become completely automatic, only requiring the user to move around the calibration object and to verify the results.

ACKNOWLEDGEMENTS

The present project is supported by the National Research Fund, Luxembourg. The authors gratefully acknowledge funding of the Erlangen Graduate School in Advanced Optical Technologies (SAOT) by the German National Science Foundation (DFG) in the framework of the excellence initiative.

REFERENCES

1. Oggier T, et al., An all-solid-state optical range camera for 3D real-time imaging with sub-centimeter depth resolution (SwissRanger), *Optical Design and Engineering*, SPIE, 5249, 534-545, 2004.
2. Svoboda T, et al., A Convenient Multi-Camera Self-Calibration for Virtual Environments, *Presence: Tele-operators and Virtual Environments*, **14**:407-422, 2005
3. Kanzow C, et al., Levenberg-Marquardt methods for constrained nonlinear equations with strong local convergence properties, *Journal of Computational and Applied Mathematics*. **172**: 375-397, 2004
4. Liu S, et al., Dynamic foot scanning using Time-of-Flight devices, *Proceedings of ISB XXIII*, Brussels, Belgium, 2011.