Maurice Lingenfelder

Design and evaluation of a point cloud recording device using dense image matching and registration in object space

Dauer der Arbeit: 5 Monate

Abschluss: April 2013

Betreuer: Dipl.-Ing. Konrad Wenzel

Prüfer: apl. Prof. Dr.-Ing. Norbert Haala

Introduction

Nowadays close-range photogrammetry is becoming a serious alternative to laser scanners in terms of accuracy, density, time and especially price. Algorithms like Semi-Global Matching enable low-cost photogrammetry to provide dense point clouds in a reasonable amount of time trying to match the quality of results obtained with laser scanners. The cameras as well as the laser scanners need to be relocated several times in order to obtain surface information from multiple views or large-scale scenes. This requires registrations for the different images or point clouds each time. One way to get sets of oriented images is the usage of Structure-from-Motion algorithms. However, those algorithms have one big limitation; they are highly reliant on feasible texture information in the images.

In this work a SLAM application called KinectFusion, using an automatic real-time ICP for sensor pose updates is described and analyzed. With KinectFusion one is able to scan whole scenes in real-time by only moving the Kinect sensor around. The thesis provides an overview of how KinectFusion works and investigates the precision of the surface reconstruction and camera orientation that can be achieved with the currently available implementation. Furthermore, the technique was used to provide the exterior orientation for an advanced configuration where two high resolution cameras enable standard stereo photogrammetry with the advantage of having pre-known camera orientations for every pose even in homogeneous textured sceneries.

KinectFusion

In 2011 Microsoft Research published a new application for the sensor Kinect, called KinectFusion. KinectFusion allows a user to freely move a standard Kinect console by hand while the algorithm estimates and visualizes the 3D surface in real-time. Thus, it represents efficient scanning of 3D objects with real time feedback. The basic idea behind the algorithm is registering new 3D information, provided as range images from the Kinect, with respect to the previous data to extract the camera orientations relative to them. A general pipeline of the algorithm includes following five main steps:

1. **Triangulation:** Depth map conversion for every range image
2. **Alignment:** Registration step, based on the iterative closest point (ICP) algorithm leading to the 6DOF of the sensor for each frame
3. **Transformation:** Calculation of truncated signed distance functions (TSDF) for every range image with known global pose

4. **Volumetric integration:** Fusing of all TSDFs into one global representation (incremental transformations to first TSDF)
5. **Raycasting:** Rendering the volume by an extraction of the isosurfaces using a per pixel raycasting method

Since KinectFusion is developped to run in realtime, these steps will iteratively repeat for every frame as soon as a new range image is generated.

SCAPE (Stereo Cameras with Automated Pose Estimation)

After KinectFusion's open sourced version was successfully built on a computer of the institute, the implementation was modified to extract the live data from the algorithm while running. Hence, the exterior orientation of the Kinect was given at frame rate.
A new system, that was designed for this thesis, consists of two industrial cameras, which are tightly mounted onto the Kinect. It combines the advantages of KinectFusion and the industrial cameras, leading to a sensor tracking and simultaneous image capturing with high resolution CCDs. The output of SCAPE is a set of image pairs with a precisely known base line and related exterior orientations. Using that information, dense image matching algorithms, for example implemented in the *BundleMatcher* developed by Konrad Wenzel, can be performed to generate dense point clouds from multiple views. As the extracted exterior orienation from KinectFusion includes errors in position, an alignment in a post processing with the single triangulated point clouds should be done.

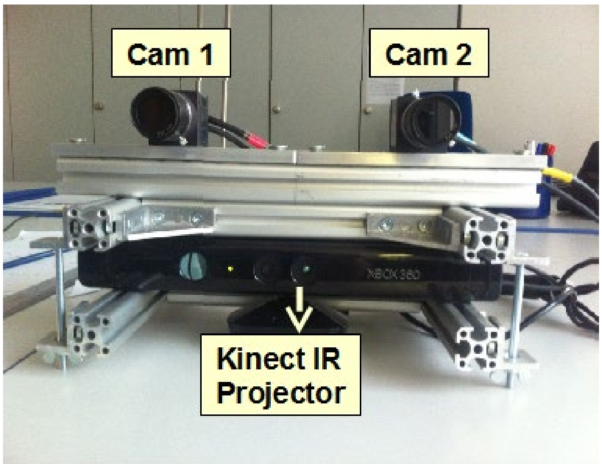

figure 1 - sensor setup

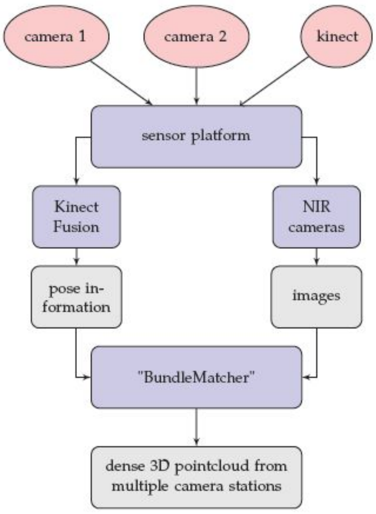

figure 2 - the idea behind SCAPE

Results

A set of test measurements was done with the depth sensor of the Kinect, with KinectFusion and by using SCAPE. All three approaches have different characteristics and therefore lead to different results in terms of best working distances, existing systematic errors, precisions or densities of the resulting point clouds. In the two tables the precisions and densities of the different approaches are presented, showing the best precisions for SCAPE in close ranges. Also, the density of SCAPE highly exceeds the one of Kinect and KinectFusion.

precisions in [mm]

	distance in [m]	0.5	0.7	1	1.2	1.5	2
Kinect	b = 7.5 cm	2.07	3.09	4.65	5.32	8.68	17.29
KinectFusion	b = 7.5 cm	0.88	0.82	1.14	1.24	1.43	2.14
SCAPE	b = 7.5 cm	0.35	0.4	0.72	1.16	-	-
	b = 16 cm	0.17	0.38	0.67	0.97	1.52	2.7

number of used points

distance in [m]	0.5	0.7	1	1.2
Kinect / KinectFusion	5'996	5'425	3'648	3'671
SCAPE	227'864	160'397	73'366	46'727

Using results of KinectFusion, investigations show differences between image space based (SIFT) and object space based (ICP) registrations. A comparison between SCAPE and (1) the SIFT-based software VisualSFM and (2) a SIFT-based solution implemented by Mohammad Abdel-Wahab was conducted. Capturing image data from two different scenes (a) rich texture & less 3D information, b) less texture and rich 3D information) shows the differences for both algorithms. Consequently, a combination with image and object space based registrations can be a powerful tool to realise a robust optical mapping system.

Conclusion

Results show that KinectFusion delivers surprisingly precise and smooth results in real-time, being able to be used in scenarios with sufficient 3D information. With SCAPE point clouds can be acquired using additional cameras in combination with a sensor pose determinated from KinectFusion. It however requires post processing but final outcomes greatly exceed those of KinectFusion in terms of density and precision. The advanced approach is capable of achieving accuracies of tenths of millimeters for short distances. Using bigger baselines, large scale mappings can be performed with reasonable precision (<4mm with 30cm base length).

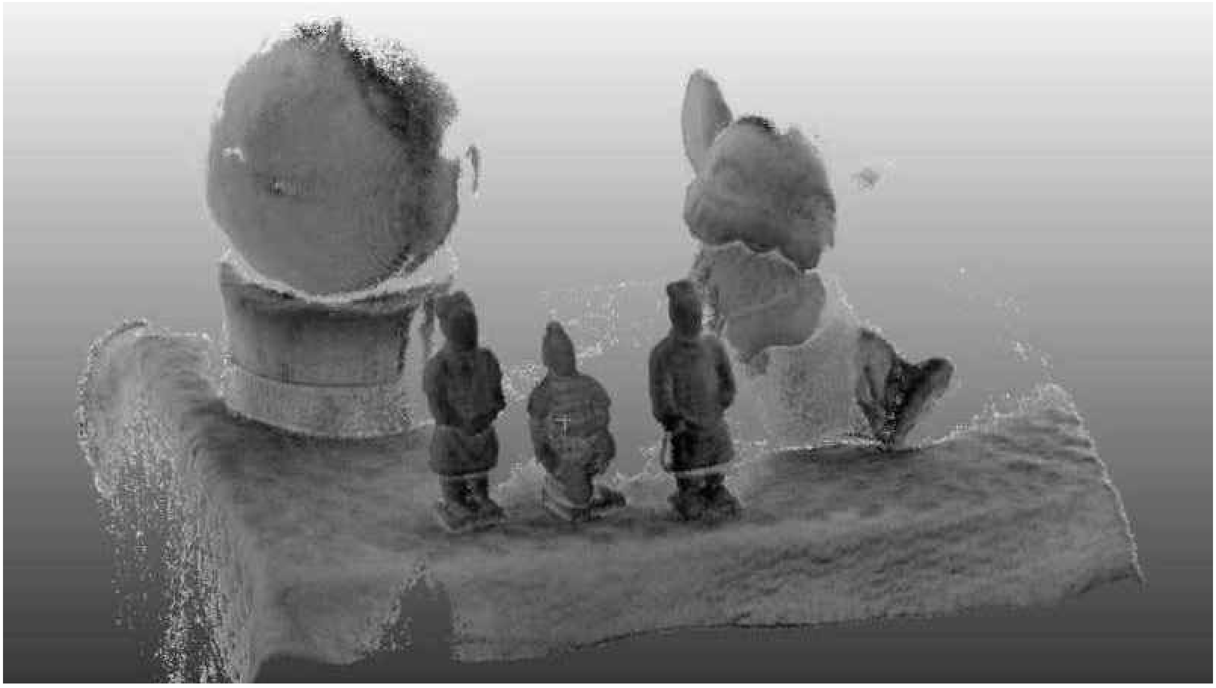

figure 3 - final point cloud determined by SCAPE from different sensor poses