our future through science

ToF camera ego-motion estimation

T RATSHIDAHO[1], JR TAPAMO[2], J CLAASSENS[1], N GOVENDER[1]
[1]CSIR Modelling and Digital Science, PO Box 395, Pretoria, South Africa, 0001
[2]School of Electrical, Electronic and Computer Engineering, University of Kwazulu-Natal, King George V Avenue, Durban, South Africa, 4041
Email: tratshidaho@csir.co.za – www.csir.co.za

ABSTRACT

We present three approaches for ego-motion estimation using Time-of-Flight (ToF) camera data. Ego-motion is defined as a process of estimating a camera's pose (position and orientation) relative to some initial pose using the camera's image sequences. Ego-motion facilitates the localisation of the robot. The ToF camera is characterised with a number of error models. Iterative Closest Point (ICP) is applied to consecutive range images of the ToF camera to estimate the relative pose transform which is used for ego-motion estimation. We implemented two variants of ICP, namely point-to-point and point-to-plane. A feature-based ego-motion approach that detects and tracks features on the amplitude images and uses their corresponding 3D points to estimate the relative transformation, is implemented. These approaches are evaluated using groundtruth data provided by a motion capture system (Vicon). The SIFT ego-motion estimation was found to perform faster when compared to the ICP-based methods.

Assisting in the localisation of autonomous robots operating in a GPS-denied environment with rough terrain, scarcity of unique landmarks and possible darkness.

INTRODUCTION

Ego-motion is an important field in mobile robotics, because it facilitates robot localisation by tracking the robot trajectory. Three algorithms for estimating the 6 degrees of freedom (DoF) ego-motion of three dimensional (3D) Time-of-Flight (ToF) cameras are evaluated and compared. A ToF camera is a compact, solid-state sensor that provides range images and amplitude images at a video frame rate of approximately 30 fps. It emits near infrared (NIR) light which illuminates the scene, and the reflected light is measured on a charge couple device (CCD) or complementary metal oxide semiconductor (CMOS) sensor of the camera. The distance is computed using the phase-shift principle.

Iterative Closest Point (ICP) is a well-known algorithm for registering two range images with partial overlap. It was concurrently introduced by Besl et al.[1] and Chen et al.[2]. Two variants of ICP namely, point-to-point and point-to-plane ICP are implemented. The third approach makes use of the amplitude images to detect and track features in the image sequence. The Scale-Invariant Feature Transform (SIFT)[3] detection algorithm is applied on the amplitude images to extract features, and their corresponding 3D points are used to estimate relative pose transform between the image sequence. An algorithm developed by Arun[4] that uses Singular Value Decomposition (SVD) is used to compute the transformation, and outliers are rejected using Random Sample and Consensus Set (RANSAC)[5].

ERROR HANDLING

The ToF camera is characterised by systematic and non-systematic errors. Systematic errors are handled by calibration. Non-systematic errors, which are also called random errors, do not have a mean value when measurements are repeated several times. They are handled by filtering. A jump edge filter is implemented in all the experiments undertaken. Jump edges occur when the transition between foreground objects and the background objects is sudden but the camera transition is smooth. The application of a jump edge is shown in **Figure 1** below.

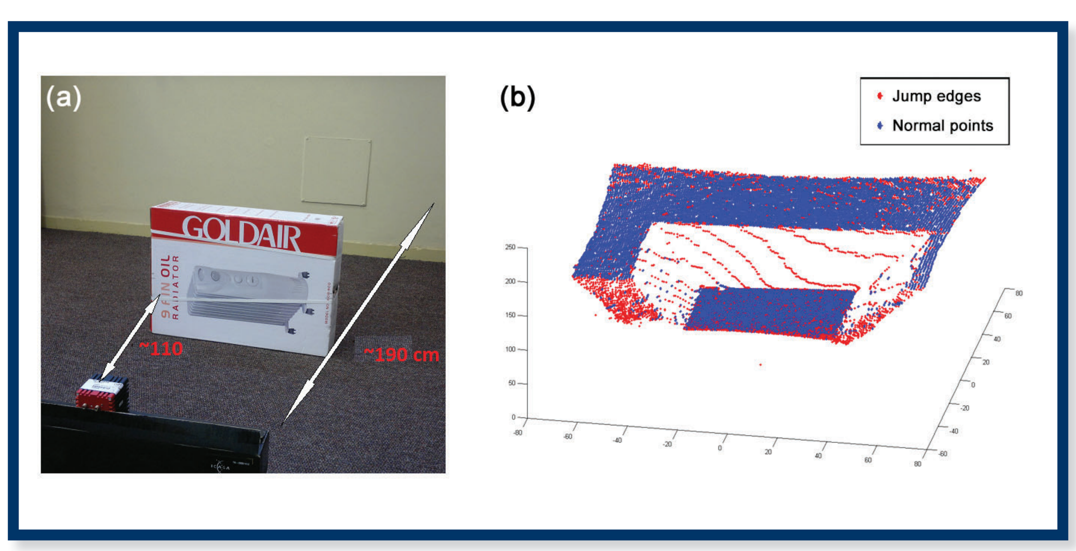

Figure 1: (a) bimodal scene used to test the jump edge filter and (b) shows the point cloud from the SR4000 ToF camera of the same scene where jump edge points are represented by red points

EGO-MOTION ESTIMATION

Point-to-point ICP

Given two 3D point clouds, a base point cloud B = $\{b_i, i=1 \ldots B_m\}$ and a scene point cloud D = $\{d_i, I=1 \ldots N_d\}$ that correspond to the same shape, the transformation (R,t) that transforms the scene point cloud to the data point cloud is computed by minimising the equation (1). This equation minimises the squared distances between the corresponding points of the base point cloud and scene point.

$$E(R,t) = \Sigma_i ||Rd_i + t - b_i||^2$$

Point-to-plane ICP

Point-to-plane ICP differs from the point-to-point in that it tries to minimise the squared distance between 3D scene points with the tangent plane to the corresponding base points. This is represented mathematically in equation (2).

$$E(R,t) = \Sigma_i [(Rd_i + t - b_i).n_i]^2$$

Feature-based Ego-motion

Image features in the amplitude images of the ToF camera are used to find the corresponding 3D points the image sequences. These correspondences are used to estimate the camera's ego-motion. SIFT feature detector algorithm is used to extract features that are invariant to rotation, scale and illumination change. SIFT features have been proven to outperform other feature descriptors based on repeatability and robustness. Transformation is estimated using a least square algorithm develop by Arun et al. [4] that uses SVD. RANSAC is used to reject outliers.

RESULTS AND CONCLUSION

Figure 2 (a) shows the results from a synthetic data experiment where a single point cloud was translated in a circular motion with a diameter of two metres. In **Figure 2 (b)** the point cloud was rotated as if the ToF camera was being rotated 360° to original starting point.

During experiments, point-to-plane converges in few iterations compared to point-to-point, but point-to-point ICP seems to produce more accurate results. This might be due to data being used. ICP fails when the transformation is purely rotational as it can be seen in **Figure 2 (b)**.

Trajectory estimation using real data captured with ToF camera is shown in **Figure 3** below. The ground truth is provided by the Vicon motion capture system. The Vicon system uses infrared reflective markers to track the pose of an object in space. It has an accuracy of sub-millimetre.

The SIFT ego-motion estimation performs best. It is faster compared to ICP based ego-motion estimation, and it produces better results.

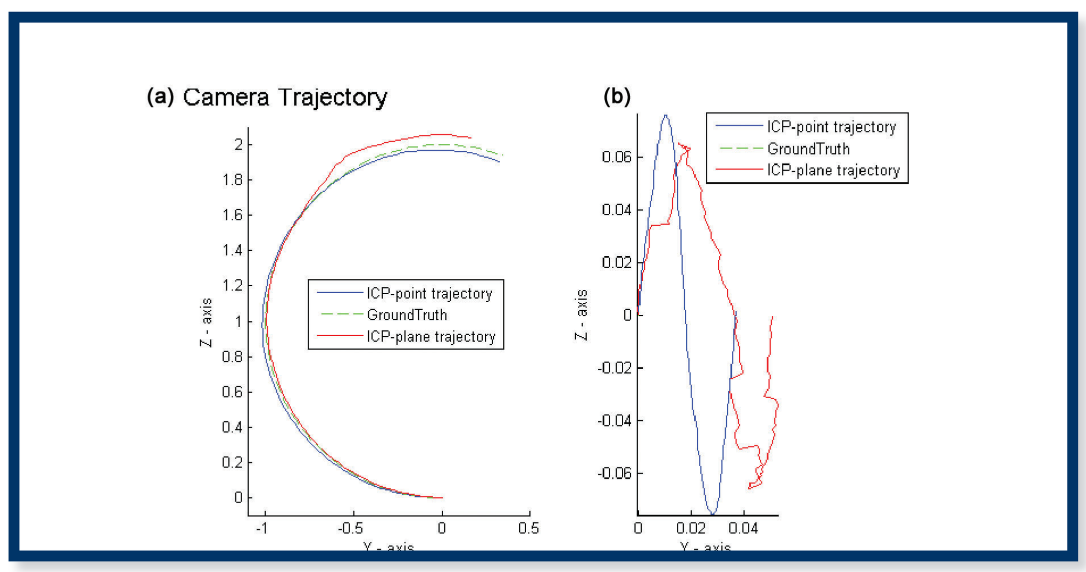

Figure 2: Trajectory estimation for (a) synthetic data test 1 and (b) synthetic data test 2. This compares the groundtruth with point-to-point ICP and point-to-plane ICP estimation

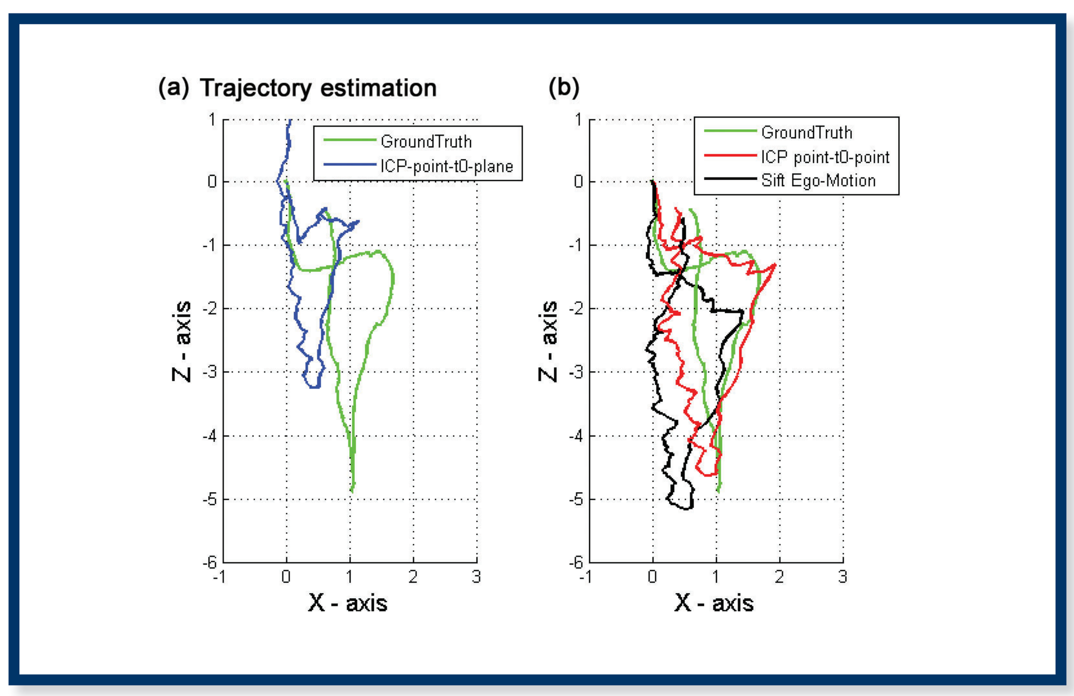

Figure 3: Trajectory estimation comparison of (a) point-to-point, SIFT and (b) point-to-plane with the groundtruth

ACKNOWLEDGEMENT

The research was funded by the Council for Scientific and Industrial Research (CSIR), Mobile Intelligent Autonomous Systems (MIAS) research area.

REFERENCES

[1] P. Besl and N. McKay. A Method for Registration of 3-D Shapes, IEEE Transactions on Pattern Analysis and Machine Intelligence, 1992.
[2] Y. Chen and G. Medioni. Object modelling by registration of multiple range images, IEEE International Conference on Robotics and Automation, 1992.
[3] D.G. Lowe. Distinctive Image Features From Scale-Invariant Keypoints, International Journal of Computer Vision, 2004.
[4] K.S. Arun , T.S. Huang and S.D. Blostein. Least-Squares Fitting of Two 3-D Point Sets, IEEE Trans. Pattern Anal. Mach. Intell., 1987.
[5] M.A. Fiscler and R.C Bolles. Random Sample Consensus: A Paradigm for Model Fitting with Application to Image Analysis and Automated Cartography, Communications of the ACM, 1981.