Intelligent Sensor-Fusion of Depth and Color Information for Cognitive Robotics

Denis Klimentjew

Technische Aspekte Multimodaler Systeme, Fachbereich Informatik
MIN-Fakultät, Universität Hamburg
klimentjew@informatik.uni-hamburg.de

1. Introduction

The presented work goes one step further than only combining data from different sensors. The corresponding points of an image and a 3D point cloud are determined through calibration. The color information is thereby assigned to every voxel in the overlapping area of the camera system and the laser range finder. Then we analyse the image and search for the locations, which are especially susceptible to errors of both sensors. Depending on the ascertained situation, we try to correct or minimize the errors. By analysing and interpreting the images as well as removing errors an intelligent tool which improves multi-sensor fusion is created. This allows us to correct the fused data and to perfect the multi-modal sensor fusion or to predict the locations where the sensor information is vague or defective.

2. Characteristics of Used Single Sensors

The advantages of the camera acquisition of images are a wide range, a higher resolution and a shorter data acquisition time. Color information is available and the costs of the images are low. The quality of image-based algorithms for depth reconstruction depends on several characteristics, e.g. lighting conditions, texture, non-homogeneous regions etc. Normally, although they yield rather good results for edges and textured areas, they fail when used for homogeneous regions. The advantages of a laser range finder are the direct acquisition of 2D points, notably without the computation overhead, an enormous amount of 2D points on surfaces, short acquisition time and independence of lighting conditions. These features render larger range finders ideal for the description of irregular surfaces. However, many laser scanners need a long “warm-up” time. This “warm-up” time is in most cases proportional to the distance. During this time the laser scanner delivers many invalid (zero) values.

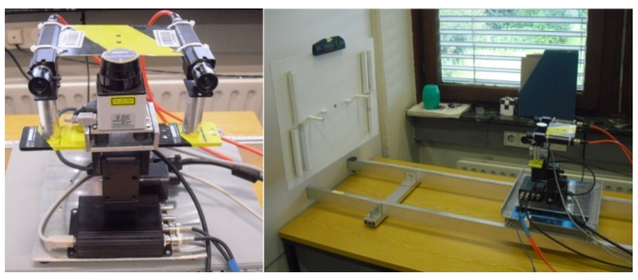

Figure 1: *The used perception platform on the left and our calibration arrangement on the right.*

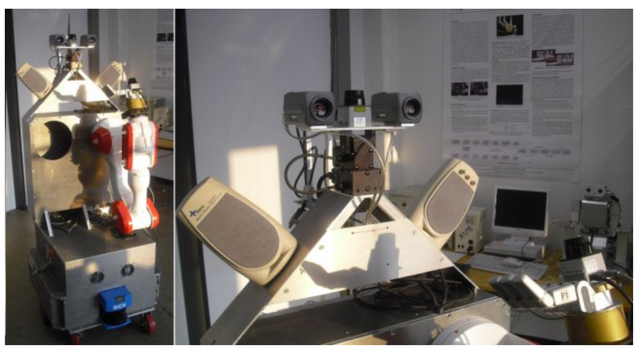

Figure 2: *The used perception platform mounted on the service robot TASER.*

After the “warm-up” time, the number of invalid values decreases, but one can still find them at the places where the reflections are large, the laser scanner receives no answer or the object lies outside the laser scanner range. For the experiments our self-developed perception platform is used. The platform consists of a 2D laser scanner, a pan-tilt unit and a stereo camera system. The perception platform is shown in fig. 1.
The 2D laser scanner together with the movable platform constitute the simulated 3D laser range finder. The setup is similar to the platform mounted on TASER, the service robot of our group. The main platform of our group is shown in fig. 2. The fact that the robot is equipped with a manipulator offers the possibility of not only recognizing objects, but also manipulating them.

3. Multi-Sensor Calibration

To create an accurate and consistent model of the environment, the scans have to be transformed into one coordinate system. This process is called registration. The result is a point cloud in the world coordinate system with the initial position (tilt position of the pan-tilt unit is 0) of the laser scanner as a coordinate origin. The process of registration is correct and is shown in fig. 3.

Figure 3: *On the left, the original image of the table scene used for our initial experiments. On the right, the reconstructed 3D point cloud.*

The result can be processed immediately, with surface reconstruction methods like Delaunay triangulation, Alpha-Shapes or the Ball-Pivoting algorithm. Fig. 4 shows the application of the Ball-Pivoting algorithm and smoothing.

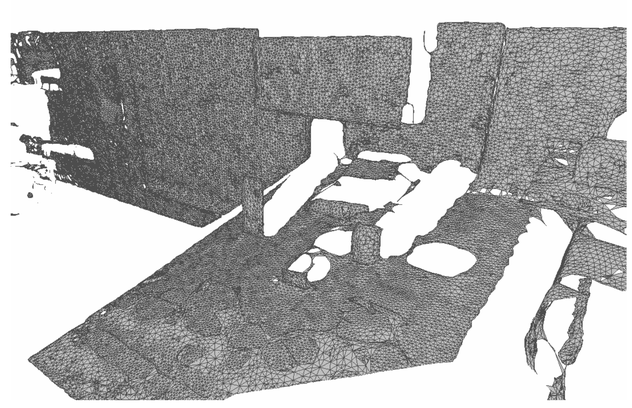

Figure 4: *The reconstructed (Ball-Pivoting) and smoothed (Laplacian smooth) point cloud.*

The calibration is based on the basic assumptions that the sensor of the laser scanner rotates on an exact circular course and thereby produces a precise line out of the points. Furthermore, we treat the data of the laser scanner as an image, which allows us to use the knowledge and the algorithms of the image processing. The fusion is possible only for the overlapping regions. Fig. 5 shows the perception areas as well as the overlapping region of both sensors schematically.

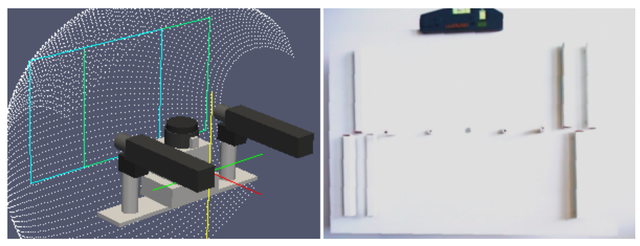

Figure 5: *The perception areas as well as the overlapping region of both sensors on the left. On the right, the used 3D calibration pattern.*

The sought transformation between laser scanner and camera is based on the extrinsic parameters of both sensors, it is independent of the arrangement of the scene or distance to the objects. The used 3D calibration pattern consists of a planar surface, two opposite so-called Y structures and some pins of different length, as is shown in fig. 5. The Y structures permit the ideal alignment of the laser scanner. Their surfaces as well as those of the pins are used later for the localization of the corresponding points for the laser scanner.
The required transformation matrix can be computed from the correspondence of image and laser scanner scene points, without requiring knowledge of the internal parameters or the relative pose of the laser scanner and camera. The transformation is determined by a numerical solution for the equations presented below.

$$\begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \begin{bmatrix} x_{lrf} \\ y_{lrf} \\ z_{lrf} \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix} \quad (1)$$

where $[r_{ij}] = f(\theta_1, \theta_2, \theta_3)$ is the rotation matrix and t the translation vector. The algorithm used here is an extension of Newton’s procedure for approximating roots on several dimensions. The procedure computes the derivations of the Jacobian matrix and approximates by means of a recursion. The inverse Jacobian matrix is determined through the Gauss algorithm. The use of the RANSAC algorithm is also possible and is implemented.

4. Intelligent Data Interpretation

As already mentioned, the resulting depth information is redundant and can be used to improve the data of the early multi-sensor fusion. Therefore we employ the depth information to improve the fusion data and use the camera images for detecting locations which are particularly susceptible to errors of both sensors. The location can be determined with standard algorithms of image processing like a Median filter or a value comparison in the HSV color space. For example, the retrieval of homogeneous regions can be accomplished with Mean-Shift-segmentation or Similarity-Measure-algorithms.

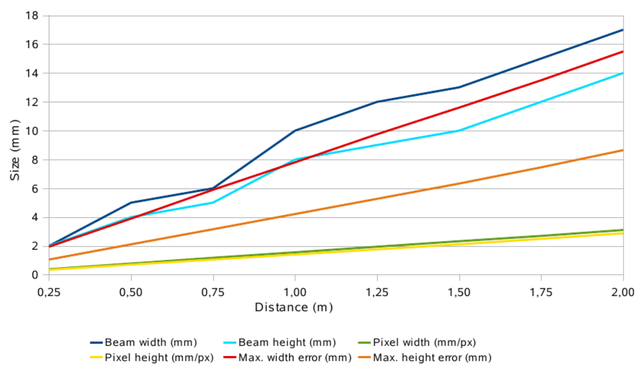

Figure 6: *The propagation of the laser beam and pixel size in relation to the distance as well as the maximal resulting error in mm.*

The following table shows the choice of the sensor for the depth information depending on the ascertained problems.

Problems	LRF	SCS
Homogeneous surfaces	×	–
Invalid (zero) values	–	×
Tiny objects	–	×
Black surfaces	–	×
Strong reflexions	–	–

Where LRF is an acronym for the laser scanner and SCS for the stereo camera system. The depth information from × marked sensors is preferred in the ascertained situation. Otherwise the more exact data of the laser scanner are used.

5. Experimental Results

The implemented application combines the early sensor fusion and interpretation components, so it allows us to achieve better results than with only the data fusion. The analysis of the images permits taking an intelligent decision for the sensor data consolidation.

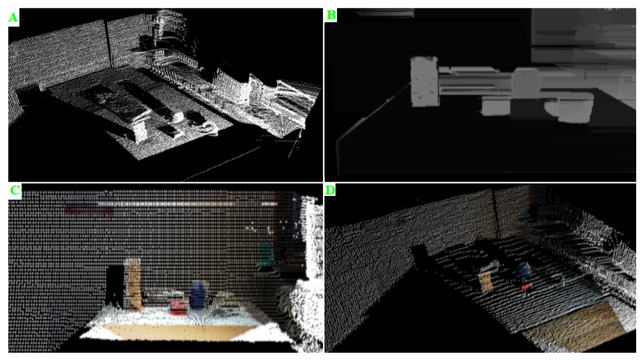

Figure 7: *a) The result of the 3D point cloud of the laser scanner. b) Disparity image of the stereo camera system. The images c) and d) show the early sensor fusion of depth and color information.*

For example, the black cup in our table scene (fig. 3) is not clearly identifiable in the 3D laser scanner image. Moreover, it is invisible after the surface reconstruction, as shown in fig. 4. The same behavior can also be observed after the early fusion, shown in images c) and d) in fig. 7. The interpretation steps deliver more suitable results, as presented in fig. 8. The representation of the cup is stable, clearly recognisable and can even be used for object recognition and/or grasp calculation.

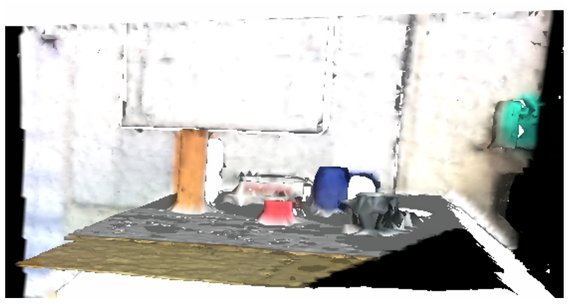

Figure 8: *The resulting reconstructed surface with the help of the Ball-pivoting algorithm from the colored point cloud includes the interpretation step. Only the fused area is presented.*

The counter-example shows the table surface. The camera finds no correspondence and therefore no depth information is available, either, as shown in image b) of fig. 7. After the interpretation the surface is clearly recognisable.