Multi Sensor Fusion of Camera and 3D Laser Range Finder for Object Recognition

Denis Klimentjew, Norman Hendrich, Jianwei Zhang

Dept. of Informatics, University of Hamburg, Vogt-Kölln-Straße 30, 22527 Hamburg, Germany
{klimentjew, hendrich, zhang}@informatik.uni-hamburg.de

Abstract— **This paper proposes multi sensor fusion based on an effective calibration method for a perception system designed for mobile robots and intended for later object recognition. The perception system consists of a camera and a three-dimensional laser range finder. The three-dimensional laser range finder is based on a two-dimensional laser scanner and a pan-tilt unit as a moving platform. The calibration permits the coalescence of the two most important sensors for three-dimensional environment perception, namely a laser scanner and a camera. Both sensors permit multi sensor fusion consisting of color and depth information. The calibration process based upon a specific calibration pattern is used to define the extrinsic parameters and calculate the transformation between a laser range finder and a camera. The found transformation assigns an exact position and the color information to each point of the surroundings. As a result, the advantages of both sensors can be combined.**

The resulting structure consists of colored unorganized point clouds. The achieved results can be visualized with OpenGL and used for surface reconstruction. This way, typical robotic tasks like object recognition, grasp calculation or handling of objects can be realized. The results of our experiments are presented in this paper.

I. Introduction

Three-dimensional reconstruction of objects and/or scenes is still a fundamental problem for computer vision research. To act in unknown environments, autonomous robots require a 3D representation of the environment. This representation depends on the aptitude of the robot's perception system to provide sufficient and accurate data. For decades, stereo camera systems were used for these tasks in robotics. The advantages of the camera acquisition of images are a wide range, a high resolution and a short 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. Here most camera-based methods achieve bad results in reconstructing the depth of continuous surfaces.

For some years now, laser range finders have been deployed for the same purpose. The advantages of a 2D laser range finder are the direct acquisition of points, notably without the computation overhead, an enormous amount of 2D points on surfaces, short acquisition time and partial independence of lighting conditions. These features render laser range finders ideal for the description of irregular surfaces. However, 3D laser scanners are still too expensive and rare, although the use of 3D laser range finders would have enormous advantages.

Both kinds of sensors, camera systems and laser scanners, offer different advantages for different tasks. Most property object recognition algorithms use shape or color information [1]. The sensor fusion of the data from a laser scanner and from cameras would extend the possibilities of robotics. This would permit the combination of the advantages of both sensors, namely, accurate distance measurement and color information. The difficulty of the fusion lies in assessing the extrinsic parameters of both sensors. These parameters describe the geometric transformations between sensor frames. Such a transformation can be identified with a calibration process. Hence, most fusion methods either need manually selected points [2][3], use the laser scanner whose trace is visible [4] or carry out the transformation from the geometrical structure. The method presented in [5] is based on the constraints between different views of a planar calibration pattern. The angle between the laser scanner line plane and the checkerboard causes an error, which is why the calibration body should be placed at a specific orientation. In [6] the authors described the fusion of 3D vision data from a stereo camera system and 2D laser scanner data. The algorithm was used successfully for the navigation, especially for environments with vertical planar walls. A comparable approach was also selected and presented in [7]. Unfortunately, many calibration methods are only partially or badly described and present no experimental results.

The rest of this paper is structured as follows. In the next section we present the system design as well as the description of the individual hardware components. Then in section III, the calibration pattern as well as the mathematical and theoretical fundamentals are described and discussed in detail. In section IV, we describe the implementation of our systems designed for a 3D colored environmental reconstruction. We discuss our experimental results in section V, and demonstrate the efficiency of the developed systems and methods. Finally, we present our conclusion and future work in section VI.

II. System design, single components and pre-processing

In this work, we use a perception platform developed by us. The platform consists of a 2D laser scanner, a pan-tilt unit and a stereo camera system. Our perception platform is shown in fig. 1.
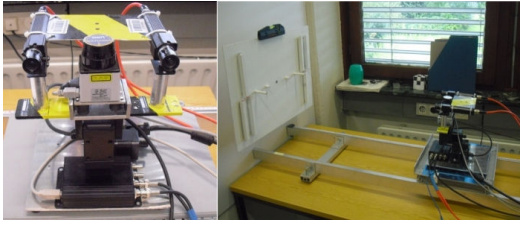
Fig. 1. The Hokuyo laser scanner and the 2 Sony cameras mounted on the pan-tilt unit. The right image shows our calibration arrangement.

We use the Hokuyo laser scanner URG-04LX and a stereo camera system consisting of 2 Sony XC-999P. The right image of fig. 1 shows our calibration arrangement, for details see section III. The stereo camera system and the laser scanner are both mounted on a pan-tilt unit. In this paper we use only one camera for the fusion of color information with the laser scanner data. 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 the TASER, the service robot of our department. The fact that the robot is equipped with a manipulator offers the possibility of not only recognizing objects, but also manipulating them. The main platform of our department is shown in fig. 2.
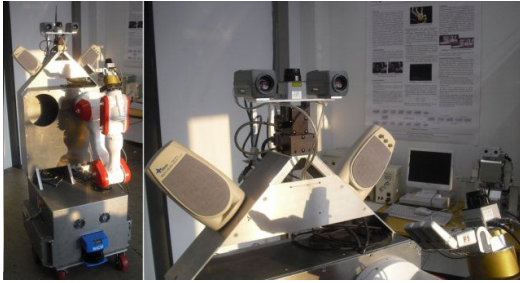

Fig. 2. Service robot TASER, the main platform of our group on the left. On the right the environment perception system consisting of two cameras and 2D laser scanner mounted on a pan-tilt unit.

A. Simulated 3D laser range finder

A movable platform and a 2D laser range finder constitute the individual components of the simulated 3D laser range finder. The used movable platform is the pan-tilt unit PTU-D46-17.5 from DirectedPerception. The pan-tilt unit is connected through serial-to-usb cable. The maximal speed is $300^o/sec$ and the maximal resolution 0.05143^o. We only use a tilt motion with the range of 77.7^o $[-46.6457^o, 31.0628^o]$ and a C/C++ interface. The mounted laser range finder is a Hokuyo URG-04LX with a SOKUIKI sensor, a compact ($50x50x70\,mm$) and light ($160\,g$) laser scanner with 240° angular range and detectable distances of $[20\,mm$ - $5600\,mm]$.

The advantage of this system is the fact that the laser scanner can still be used in 2D without any physical changes. The point density depends on the physical properties of the single components and is shown in fig. 3. We examined three different Hokuyo URG-04LX laser range finders and tested
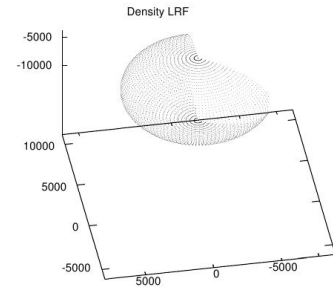

Fig. 3. Point density for simulated 3D laser range finder based on the 2D laser scanner and pan-tilt unit.

them under different conditions. The important awareness for this work was, that all distances to black surfaces are clearly shorter than in reality. As a result, this color cannot be used for the construction of the calibration pattern. In addition, dark velvets and reflective surfaces strongly falsify the results.

B. Pre-processing tasks

One of the first assignments was the handling of a number of invalid (zero) values, to compensate for the incorrect information or for those places where no value was ready. To compensate for these errors we use an interpolation procedure inspired by Shepard and presented in eq. 1. At first we compensate only over the horizontal neighbors, because we fuse the data from the laser scanner and the camera line by line. In addition, in the resulting 3D point cloud we weight the horizontal neighbors more than the vertical ones, because the laser range finder beam extends more in the horizontal than in the vertical direction.

$$a_{zi} = \frac{\sum_{j=1}^{n} z_j * d_j^{-p}}{\sum_{j=1}^{n} d_j^{-p}} \qquad (1)$$

where a_{zi} is an average for point i, p is an exponent of the distance, d_j is a Euclidean distance to n neighbors and z_j is a value of n next neighbors.

The resolution of our system is limited in the horizontal direction by the laser scanners and the camera. In the vertical direction the resolution can be changed according to demand. This depends only on the resolution of the movement platform, see section I. A plane with 682 data points is scanned in $100\,ms$ by the 2D laser range finder, a rotating mirror device. Scanning the environment with a laser scanner and a pan-tilt unit is done in a stop-scan strategy. To recognize the 3D surroundings, the position of the moving platform is continuously changed. One of the reasons for this platform construction is the reduction of the parallax effect.

$$\begin{bmatrix} c(\varphi)c(\theta) & -s(\theta)c(\varphi) & s(\varphi) & -d_z s(\frac{\varphi}{2}) \\ s(\theta) & c(\theta) & 0 & 0 \\ -c(\theta)s(\varphi) & s(\theta)s(\varphi) & c(\varphi) & -d_z s(\frac{\varphi}{2})c(\frac{\varphi}{2}) \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (2)$$

where c and s are cos and sin respectively, θ is the deflection angle of the laser beams (multiple of angular resolution) and φ is the current angle of the pan-tilt unit, d_z is a vertical vector to the coordinate origin.

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. We define the coordinate origin on the ground under the middle of the laser range finder sensor and use a right-handed 3D coordinate system for both systems. This choice of the coordinate origin later permits an easier portability to the robot's coordinate system used for our service robot TASER. The $[x, y, z]$ positions depend on the physical structure and are calculated through trigonometrical functions and rotation around the y-axis. We use only the pitching direction, so only the tilt unit of the pan-tilt unit is used. The resulting matrix is presented in eq. 2. After multiplication with the array of distances delivered by the laser scanner (single array element $[data[n]_{LRF}, 0, 0, 1]^T$), the following structure results (eq. 3), which can be implemented simply and needs only minor computing time.

$$\begin{bmatrix} x_n \\ y_n \\ z_n \\ 1 \end{bmatrix} = \begin{bmatrix} data[n]_{LRF}cos(\varphi)cos(\theta) - d_z sin(\frac{\varphi}{2}) \\ data[n]_{LRF}sin(\theta) \\ data[n]_{LRF}cos(\theta)sin(\varphi) - d_z sin(\frac{\varphi}{2})cos(\frac{\varphi}{2}) \\ 1 \end{bmatrix} \quad (3)$$

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. 4.
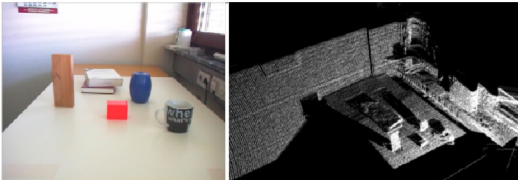

Fig. 4. On the left, the original image of the table scene used for our initial experiments. On the right, the reconstructed 3D point cloud (approx. 250,000 voxels). The perspective is changed to give a better view.

The result can be processed immediately without any changes, for example, with surface reconstruction methods, like Delaunay triangulation, Alpha-Shapes or the Ball-Pivoting algorithm. Fig. 5 and fig. 6 show the application of the Ball-Pivoting algorithm with different radii and smoothing levels.

The next pre-processing step is the calibration of the single cameras. First of all, the intrinsic calibration parameters of the camera system are identified. Then the lens distortions of the input images are calculated. We use our own implementation, based on the algorithm by Z. Zhang,
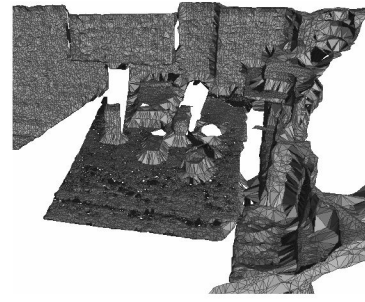

Fig. 5. The reconstructed point cloud with the Ball-Pivoting algorithm. The rectangle standing out from the wall is the calibration body, see fig. 1.

which was introduced in detail in [9]. The fully automatic procedure is fast, stable and easy to use.
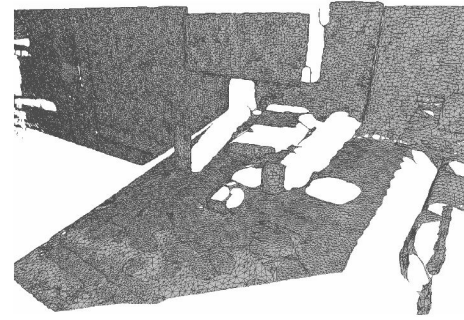

Fig. 6. The reconstructed (Ball-Pivoting) and smoothed (Laplacian smooth) point cloud. The rectangle standing out from the wall is the calibration body, see fig. 1 and 8.

The pre-processing is concluded with the smoothing of the images by a Gauss filter to minimize the noise. Therefore the single sensor data are prepared for the fusion.

III. MULTI-SENSOR CALIBRATION

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. Both kinds of sensors have different characteristics and perception areas. The fusion is possible only for the overlapping regions. Fig. 7 shows the perception areas as well as the overlapping region of both sensors.
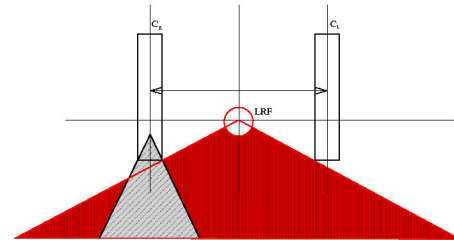

Fig. 7. The perception areas as well as the overlapping region of camera and laser range finder.

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 original idea was to combine the calibration patterns for the camera and the laser scanner. Basically we use the typically checkered pattern which is extended with a 3D structure. Because of the difficulties of the Hokuyo URG-04LX laser scanner with the black surfaces, as already mentioned in section II, the checkered pattern was renounced in the design of the calibration pattern.

The 3D calibration pattern consists of a planar surface, two opposite so-called Y structures and some pins of different length. 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. We use the color information for the localization of corresponding pixels in the image (red lines at the Y-structure and points at the front of the pins).

Fig. 8. The planned calibration pattern on the left and the 3D calibration pattern resulting from several experiments on the right.

The resulting 3D calibration body is shown in fig. 8, the lateral view of the calibration body can be seen in fig. 1. The difficulty was to design the 3D surfaces so that they differed from the planar body and from each other, but would not be recognized by the laser scanner as outliers and be removed. The design was adapted in the course of the experiments. The best calibration results were reached with a distance to the calibration pattern of approx. $0.70\,\mathrm{m}$ - $1.00\,\mathrm{m}$.

The calibration runs in the following steps. First the laser scanner is moved by the pan-tilt unit and scans the environment in coarse steps (0.25^o - 0.5^o). Through changes in the depth information, our system localizes the region of interest (ROI) and the changeover from one to two teeth (or vice versa) in the Y-structure and rescans it in finer steps, up to 0.05143^o, to find the desired position. The left image in fig. 9 visualizes three successive laser scans.
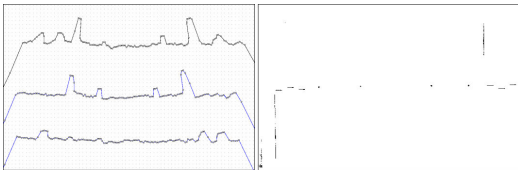

Fig. 9. The left image shows the gradual passing of the pan-tilt unit from the top to the bottom of the calibration body consisting of pins and both Y structures. The top dataset shows two amplitudes on the left and one on the right, signifying the respective teeth of the Y-structures. In the middle dataset, only the amplitudes of the pins are visible. In the lowest set, the laser has passed the middle of the calibration body. Consequently, one amplitude is now visible on the right and two on the left. In the right image the color threshold value is applied to the camera image.

When the laser scanner has detected the characteristic

pattern for the correct position in relation to the calibration body, the data of the laser scanner corresponds to the top line of the left image in fig. 9 and the pan-tilt unit moves to that position. After this the platform is stopped and a camera image is acquired. The right image of fig. 9 shows the camera image after applying the color threshold value. The Y-structures and the pins are clearly visible in the laser scan and the camera image. Thus we have the coordinates of three teeth with two points in each case (beginning and end) as well as four single pins, see images in the figure above. Through the found points of both images it is possible to calculate the straight line for each case. The correspondences between beginning and end of both lines are unambiguously fixed by the Y-structure. Because the pixels and voxels are also clearly identified in the line, the corresponding points can be associated in corresponding pairs. Consequently, with this number of corresponding points we have an overdetermined model. Moreover, the rotation angle between two lines can be calculated and both images rectified in relation to each other.

The required transformation matrix between a laser range finder and a camera is independent of scene structure. However, it can be computed from the correspondence of image and laser scanner scene points, without requiring knowledge of the internal parameters or relative pose of the laser scanner and camera. The transformation is determined by a numerical solution for the equations presented below. The transformation is defined by three rotations $\theta_1, \theta_2, \theta_3$ and three translation components t_1, t_2, t_3. The rotation matrix R and the translation vector t are searched.

$$\begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \begin{bmatrix} x_{lrf} \\ y_{lrf} \\ z_{lrf} \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix} \quad (4)$$

where $[r_{ij}] = f(\theta_1, \theta_2, \theta_3)$ is the rotation matrix and t the translation vector. The matrix presents a transformation from a 3D laser scanner to 2D image data.

The algorithm used here is an extension of Newton's procedure for approximating roots on several dimensions. Very important for the presented results is the approximation of start values which can be simply met for all trigonometrical functions of rotation angles as an interval $[-1, 1]$. 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. The resulting matrix is similar to the matrix calculated with the extension of Newton's procedure.

As already mentioned above and illustrated in fig. 7, the perception areas of the two sensors do not precisely correspond. The perception range of the camera is smaller, which has two consequences. First we receive the color information only for the laser scans, i.e. only for the overlapping area of both sensors. The matrix in the eq. 4 is determined up to the scaling factor. As a result, the overlapping area must be newly calculated for each pair of corresponding

scans. Second, the transformation matrix can be used in the overlapping area only, otherwise the transformation would cause an error and produce wrong correspondences.
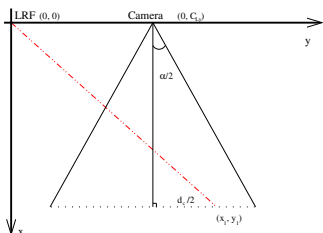

Fig. 10.   Geometrical relation between laser scanner and camera.

For the computation of the overlapping area we use elementary geometry, which is visualized in fig. 10. With the help of the law of sines the length of the distance d_c can be calculated, namely

$$d_c = 2x_1 \frac{sin(\frac{\alpha}{2})}{sin(90^0 - \frac{\alpha}{2})} \quad (5)$$

To decide now whether a point lies in the overlapping area, only the following condition must be checked:

$$y_1 \in [\ C_{Ly} - \frac{d_c}{2};\ C_{Ly} + \frac{d_c}{2}\ ] \quad (6)$$

The single coordinates of the points result from eq. 4, which determines 3D world coordinates. The angle α is the flare angle of the used camera. Last of the determining components is the distance between both sensors, translation vector C_{Ly} in the fig. 10. For the purpose of exact measurement we use the corresponding points determined before, and calculate the fundamental matrix F, presented in eq. 7, with the Eight-point algorithm.

$$\begin{bmatrix} -0.00610919 & -0.06188764 & -1.40267897 \\ -0.26569065 & -3.27804565 & 0.72559673 \\ 78.98945618 & 972.49920654 & 1.00000000 \end{bmatrix} \quad (7)$$

The red number in the fundamental matrix (eq. 7) is the request distance in mm. If the parameters have been determined, the overlapping area has been calculated and the transformation is unambiguous, the multi-sensor fusion can be accomplished.

IV. IMPLEMENTATION

The work-flows of the three-dimensional reconstruction systems developed in this paper are introduced in fig. 11 depending on the pan-tilt unit solution. For our initial experiments we use the same table scene, see the left image in fig. 4. The right image shows the unorganized point cloud reconstructed with the simulated 3D laser scanner. The perspective is changed to give a better view.

The fusion takes place in the early phase and occurs line by line. First, the position of the laser scanner is continuously changed by the pan-tilt unit, and for each step the laser data and camera image are acquired. In relation to
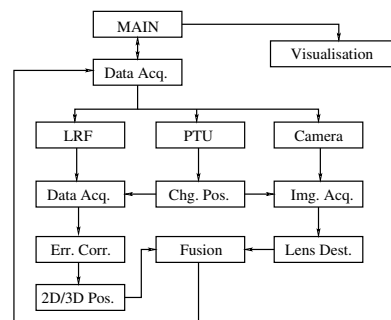

Fig. 11.   Simplified flow-chart of our 3D colored reconstruction system.

the camera vertical flare angle we use a step size of 0.07638^o to get a match of the number of lines in the image and laser scan.

Further processing is carried out separately. The lens distortion is calculated from the image, see section II or our description in [9] for more details. The laser scanner data are the distances to the obstacles in the surroundings. After this, the invalid (zero) values, see section II, are corrected by the described windowing method and the data saved in the simple data structure. The positions of single voxels are calculated by multiplication with the registration matrix, see eq. 2-3 in section II. Then we rectify the image in relation to the laser scanner line with parameters obtained from the fundamental matrix. The corresponding line and the relation between the voxels of the laser scanner and the camera pixels are calculated with the help of mathematical functions. All required parameters like the positions of the single laser beams, the above-mentioned relation or the camera flare angle are known. The positions of single voxels and corresponding camera pixels are determined by multiplication with the matrix, as shown in section III. For the acceleration of the calculation we determine the border of the overlapping area of both sensors first, associate the color information to every voxel of the laser scanner line and save the resulting structure. In the end, all lines are combined into an image-like structure, 3D coordinates and the color information are included.

The system adapts itself dynamically, so that the number of the points can vary depending on the user demands and settings of the moving platform.

V. EXPERIMENTAL RESULTS

For the demonstration of our results we use the table scene presented in section II, see fig. 4. We present the results of the multi-sensor fusion between a laser range finder and a camera system mounted on a pan-tilt unit in fig. 12.

The results are partially colored point clouds with $249,216$ voxels. The perspective was changed for the purpose of a better view. The view area of the laser range finder is limited in respect to the construction and amounts to around 124.1^o. The flare angle of the camera is a 57.48^o horizontal and 44.06^o vertical angle with a resolution of 704 x 576 pixels respectively, the resolution of the pan-tilt unit can be adapted. Therefore, the maximum view area of the
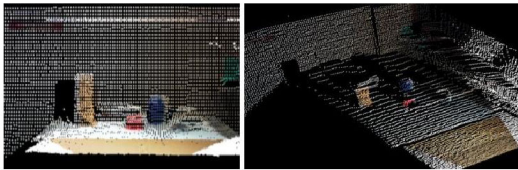
Fig. 12. The result of the multi-sensor fusion between a laser range finder and a camera system: colored point cloud. Two different perspectives are presented for the purpose of a better view.

Fig. 14. The propagation of the laser beam and pixel size in dependence to the distance as well as the maximal error resulting from it in *mm*.

complete system amounts to 124.1^o x 77.7^o, colored 57.48^o x 44.06^o. The resulting algorithm is implemented in C/C++ and visualized with OpenGL. All presented examples have a view area of 124.1^o x 44.06^o.

We use the Ball-pivoting algorithm again for surface reconstruction, this time for the colored point cloud. The result can be seen in fig. 13. Only the fused area is shown. The color for the recumbent voxels outside the fused area is set to black.

Fig. 13. The resulting reconstructed surface with the help of the Ball-pivoting algorithm from the colored point cloud. Only the fused area is presented.

The developed systems inherit the errors and properties of their individual components, therefore the results of the multi-sensor fusion are difficult to evaluate. The size of the laser beams increases with the distance, besides, the shortest measured distance or an average value of the measured distances will be delivered. At the same time the size of the surface references to an image pixel increases, however, the image pixel still has only one value per color channel. Another source of errors is the Parallax effect which increases with the distance between both sensors. An error is thereby produced the minimum of which lies in the middle area and increases outwardly. In contrast, the effect of Parallax decreases with an increasing distance.

During the empiric experiments we assess a maximal error of approximately 5 pixels in the horizontal and 3 pixels in the vertical direction respectively. Besides, several table scenes with differently placed objects were examined. The propagation of the laser beam and pixel size in dependence to the distance as well as the maximal error resulting from it in *mm* are summarized in fig. 14.

The temporal performance of the whole system is linear. The time performance is directly proportional to the resolution of the pan-tilt unit and the number of the 3D points and is not yet optimized completely.
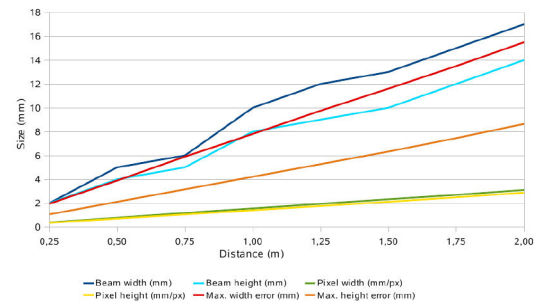
VI. CONCLUSION AND FUTURE WORK

In this paper, we proposed a perception system which permits the fusion of a 3D laser scanner and a camera. The result is a 3D colored point cloud. The presented perception system opens up enormous possibilities for robotics. The results are thoroughly comparable with the application of the same single algorithms to the data of separate sensors.

The method is interesting first of all for object recognition, but also for the other robotics areas. The combination of a precise distance measurement and color information permits the application of several methods and data fusion on a high level.

One of the next steps is an attempt to fuse two similar kinds of sensor data, the input from the 3D laser scanner and the stereo camera system.

VII. ACKNOWLEDGMENTS

REFERENCES

[1] N. Osaka, I. Rentschler, and I. Biederman, Object Recognition, Attention, and Action, ISBN 978-4431730187, Springer-Verlag, 2007.
[2] D. Scaramuzza, A. Harati, R. Siegwart, Extrinsic Self Calibration of a Camera and a 3D Laser Range Finder from Natural Scenes, International Conference on Intelligent Robots and Systems, USA, 2007.
[3] C. Altuntas, F. Yildiz, Registration of Terrestrial Laser Scanner Point Clouds by One Image, The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. Vol. XXXVII. Part B5. Beijing 2008.
[4] J. Forest, J. Salvi, A Review of Laser Scanning Three-dimensional Digitizers, In IEEE/RSJ International Conference on Intelligent Robots and System. EPFL Lausanne, Switzerland, pp. 73-78, 2002.
[5] Q. L. Zhang, R. Pless, Extrinsic Calibration of A Camera and Laser Range Finder (Improves Camera Calibration), IEEE/RSJ International Conference on Intelligent Robots and systems, Sendai, Japan, pp. 2301-2306, 2004.
[6] H. Baltzakis, A. Argyros, and P. Trahanias, Fusion of laser and visual data for robot motion planning and collision avoidance, Machine Vision and Applications, 15:92-100, 2003.
[7] S. Wasielewski, O. Strauss, Calibration of a multi-sensor system laser rangefinder/camera, Intelligent Vehicles Symposium., pp. 472-477, ISBN: 0-7803-2983-X, USA, Current Version Published: 2002.
[8] D. Klimentjew, M. Arli, J. Zhang, Three-Dimensional Scene Reconstruction Based on a Moving Two-Dimensional Laser Range Finder for Service-Robots, IEEE International Conference on Robotics and Biomimetics, 2009.
[9] D. Klimentjew, A. Stroh, S. Jockel, J. Zhang, Real-Time 3D Environment Perception: An Application for Small Humanoid Robots, IEEE International Conference on Robotics and Biomimetics, Thailand, pp. 354-359, 2009.