Robust Detection and 3D Reconstruction of Boundary Segmented Objects in a Robotic Library Scenario

S. Natarajan[1], S.M. Grigorescu[2], D. Mronga[3], A. Gräser[1]
[1]Institute of Automation, University of Bremen
[2]Institute of Automation, Transylvania University of Braşov
[3]German Research Center for Artificial Intelligence, University of Bremen
{natarajan, ag}@iat.uni-bremen.de, s.grigorescu@unitbv.ro, dennis.mronga@dfki.de

Keywords: *Robust feature extraction*, *Pose estimation*, *3D Reconstruction*, *Feedback control in image processing*.

Abstract

This article investigates the problem of robust feature extraction and 3D reconstruction of boundary segmented objects. The considered objects are firstly segmented using the Canny edge detector and further detected with the help of the Hough transform. The robustness of object boundary detection is achieved through the inclusion of a cascade feedback control system at object recognition level. An accurate and efficient pose estimation algorithm is presented with the purpose to estimate the location of three dimensional objects from the extracted 2D features. The proposed method needs no a-priori knowledge about the location and appearance of the objects of interest. The advantage of the presented system is that object recognition is independent of illumination conditions and functions reliably in clustered environments. For performance evaluation, the approach has been tested within the library scenario of the rehabilitation robotic system FRIEND.

1. Introduction

In autonomous service robots, the stereo vision system represents the way in which a robot perceives the outside world. It enables to robot to acquire the 3D depth information which is vital for recognition and manipulation of objects in an unstructured environment. FRIEND (*Functional Robot with dexterous arm and user-frIENdly interface for Disable people)* is a semi-autonomous rehabilitation robotic system designed to increase the autonomy of elderly and disabled people in their daily and professional life activities [1]. It comprises of a dextrous manipulator arm with 7 *Degree-of-Freedom* (DoF) mounted on an electrical wheel chair. A Bumblebee® stereo vision camera is attached to a 2-DoF pan-tilt head unit mounted on a rack behind the user. The camera enables the view of the scene in front of the user, including the platform and the manipulator fixed, as seen in Figure 1(a). The machine vision framework of the FRIEND system is known as ROVIS (*RObust machine VIsion for service robots*) [2].

The main objective of the library scenario is to integrate disabled persons back into professional life through a working support scenario which treats the handling of books at a library desk. The robot's manipulator is used to autonomously grasp books and bring them to the user for the purpose of lending them. In order for the robot to dextrously grasp a book, the vision system has to robustly recognize the books in the 2D image followed by a reliable 3D reconstruction. Since books come in different sizes, colours and with varying texture on the cover, the object recognition algorithm must be robust with respect to handling such scene uncertainties. A typical scene from the library scenario is illustrated in Figure 1(b).

This paper is organized as follows. Section II presents the state of the art of object recognition and 3D reconstruction. In Section III, the ROVIS object recognition approach is detailed, whereas the 3D Reconstruction algorithm is described in Section IV. In Section V, the obtained experimental results of the proposed system are presented. Conclusion and future work are described in Section VI.

(a) (b)

Figure 1: (a) The rehabilitation robot FRIEND operating in a library. (b) A library desk scene imaged by the FRIEND's global camera.

2. Related Work

Within the autonomous robot librarian presented in [3], the system uses an camera in hand approach to recognize labels of books using *optical character recognition* (OCR) methods and a special gripper to manipulate books from the shelf. In contrast to the existing system, our scenario has to perform object recognition in clustered environments and the 3D reconstruction accuracy must be precise enough in order to grasp books with a standard gripper. A SIFT model based recognition system which uses the object's local texture is described in [4]. The method can extract full pose estimation of the textured objects. Since books with low texture, thus low pattern information, can also exist in the scene, SIFT based recognition will not provide a suitable solution.

Various triangulation methods are available in literature to reconstruct the 3D points from corresponding image points. Beardsley *et al.* [5] proposed 'quasi-Euclidean' based reconstruction, where an approximation to the correct Euclidean frame is selected and then the shortest perpendicular line segment which joins the projection lines of the two image points is computed. For the nonlinear method in [5], the 3D reconstruction accuracy is low when compared with linear methods. The polynomial method proposed in [6] claims high reconstruction accuracy under the assumption of Gaussian noise model. Since the global minimum of the sum of magnitude of image errors has to be calculated, the computation cost of the algorithm is high and thus not suitable for the FRIEND system. The POSIT algorithm proposed in [7] provides an iterative solution to the pose estimation problem, but also requires prior knowledge of the 3D model of the object. This imposes restriction in reconstruction of arbitrary objects in the unstructured environment which is contrary to our scenario.

The main contributions of this article are the inclusion of feedback control at boundary extraction level with the purpose to improve the robustness of the overall object recognition system. Also, an optimal stereo triangulation algorithm, along with plane based object orientation estimation, will be presented. Its goal is to obtain an object's pose relative to the reference coordinate system of the robot, also known as the *World Coordinate System*.

3. Object Recognition through Boundary Object Segmentation

The first step in the object recognition chain is the definition of the *Image Region of Interest* (ROI) containing the objects to be recognized. This is performed by detecting a marker placed on the library desk [2]. The ROI definition minimizes the object search area in the 2D stereo image and hence the computation time. Since the general shape of the book is invariant, it can be described by straight lines, that is, its four corners which will represent the object feature points. The blue circles in Figure 3(c) represent the extracted feature points. As convention, the first book corner is assumed to be the top-left one and the numbering of the feature points is made in a clockwise manner. The four feature points are defined as:

$$p_{Li} = (x_{Li}, y_{Li}), \\ p_{Ri} = (x_{Ri}, y_{Ri}), i = 1,2,3,4. \quad (1)$$

where p_{Li} and p_{Ri} represent book corners in 2D image coordinates (x_i, y_i) in left and right images, respectively.

A. Open Loop Object Recognition

The basic structure of the image processing chain implemented in ROVIS for recognition and localization of all the books in the left and right stereo images is shown in Figure 2. Boundary based edge segmentation is performed using the Canny edge detector [8] which detects sharp local changes in the intensity of an image. Canny lower threshold T_L and upper threshold T_H parameters are used to detect the strong and weak edges along the object boundaries, as seen in Figure 3(c). The low threshold can be expressed as the function of high threshold as T_L = 0.4 T_H. Due to the presence of image noise and non uniform illumination, the contour edges obtained from the edge detector may not be continuous and have discontinuities in the intensity image [9].

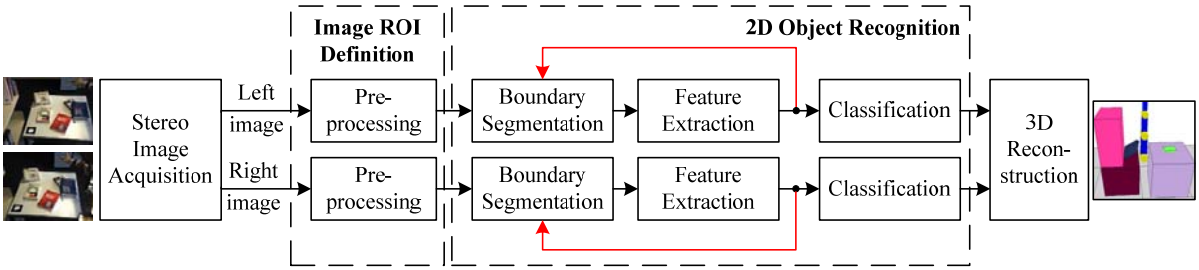


Figure 2: Boundary segmented object recognition and 3D reconstruction chain in ROVIS.

Hough transform [10] is then applied on the boundary segmented image to extract the line features from the edge pixels as it is robust to the image noise and can cope with discontinuities in feature boundaries. The binary edge pixels are mapped into the so-called accumulator array which gets incremented when foreground pixels lie on a straight line. The entries in the accumulator above a threshold value T_{HG} yields the information about the location of the most significant straight lines in the image space. The corners of the book cover are computed by the intersection of the computed Hough lines, illustrated by red lines in Figure 3(c).

The main drawback of open-loop object recognition is the usage of fixed image processing parameters which cannot adapt to the varying imaging conditions. This phenomenon is illustrated in Figure 3, where images of the same scene acquired under different illumination conditions are shown. Constant image segmentation parameters obtained from artificial lighting conditions are used for segmenting both sets of images.

Although feature extraction is reliable in the case of artificial illumination conditions, object recognition fails when the same parameters are used for scenes imaged in daylight conditions. Hence, the selection of optimal thresholding parameters T_H and T_{HG} is crucial for robust object recognition under varying illumination conditions.

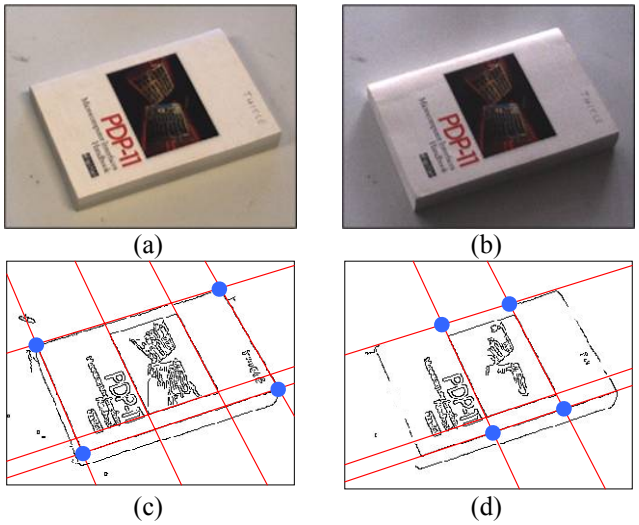


Figure 3: Image of the same scene acquired under artificial (a) and daylight (b) illumination conditions. (c) and (d) object feature points extraction using constant boundary segmentation parameters.

B. Feedback Control of Boundary Recognition

In order to obtain good object recognition results independent of illumination conditions clustered environments, we propose a closed-loop cascade control structure which regulates the parameters of boundary image segmentation, as shown in Figure 4(a). The control objective is to maximize object recognition results via the maximization of the control variable y.

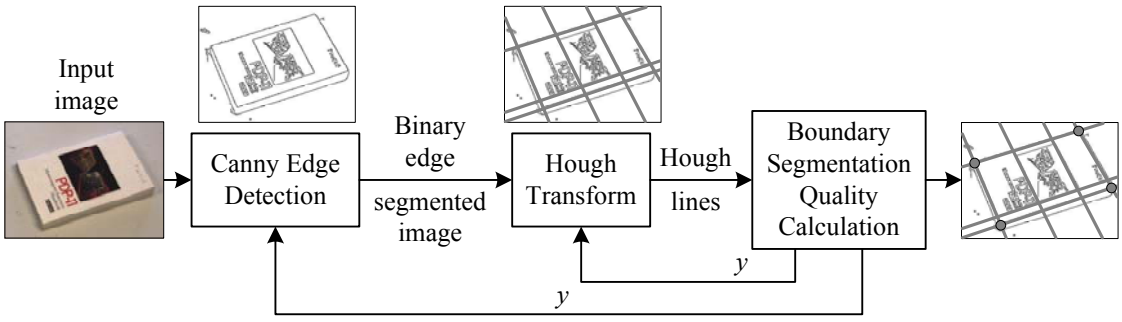


Figure 4: Cascade control structure for robust boundary segmentation.

The measure of the boundary segmentation quality calculated at image feature extraction level is used as a controlled variable to control the segmentation parameters T_H and T_{HG} in the proposed closed-loop system. The equation of the proposed measure is:

$$y = \begin{cases} e^{N/N_{\max}} \cdot \sum_{n=0}^{N_\#} \dfrac{N_f(n)}{A_{ROI}}, & \text{if } N \le N_{\max}, \\ 0, & \text{if } N > N_{\max}, \end{cases} \tag{2}$$

where $N_\#$ represents the number of candidate solutions and $N_f(n)$ the number of foreground pixels covered by the detected lines of the n^{th} object, normalized with the area of the image

ROI, A_{ROI}. To reduce the computational time of the Hough transform, the maximum number of lines allowed in an image ROI is set to a constant value N_{max}. The exponential term in Equation 2 enhances the detection of objects with minimal number of lines and the summation term enhances the quality of the detected objects. Optimal determined values of the Canny and Hough transform thresholds ensures that a reliable input is given to the feature extraction module which directly influences the accuracy of 3D reconstruction. The following section will detail the object 3D reconstruction using the obtained robust 2D feature points.

4. 3D Reconstruction

In order to dexterously grasp and manipulate an object using a robot manipulator, the object's pose and its 3D model has to be precisely computed. The pose estimation, which involves obtaining the object's 3D location and orientation, is performed in the 3D reconstruction module using the extracted 2D feature points from the left and right stereo images, as shown in Figure 2. The first step of object reconstruction is to calculate the 3D position of the four extracted book corner points using an optimal stereo triangulation algorithm. A novel object orientation estimation method based on the reconstructed 3D object points will be detailed in Section 4.B. The reconstruction of one 3D book corner point is described below. The same procedure is used to reconstruct the other three book corner points.

A. Optimal Stereo Triangulation

Let $P(x, y, z, 1)$ be the considered object 3D point which is projected into 2D image feature points as p_L and p_R in the left and right stereo images respectively. The real world 3D coordinates are related to image coordinates using the left and right camera projection matrices Q_L and Q_R, respectively. These matrices describe the homogenous transformation between the robot's reference coordinate system W, located at the base of the manipulator arm, and the left C_L and right C_R coordinate systems of the stereo camera, respectively. The projection matrices are computed and initialized using the ROVIS Camera Calibration procedure [12]. We consider a linear mapping from 3D points to 2D image coordinates in both left and right stereo images on homogenous coordinates. The relationship between the considered 3D point P and its perspective 2D image projections from Equation 1 is given as:

$$\begin{aligned} p_L &= Q_L \cdot P, \\ p_R &= Q_R \cdot P, \end{aligned} \tag{3}$$

With the known camera calibration parameters and robust 2D image points, the two projection lines in the 3D space corresponding to the image points can be computed. This computation is formulated as triangulation problem [13]. In absence of noise, trivial solution exists for the triangulation problem. Since the stereo images acquired at low illumination conditions have more noise, the two projective rays will not intersect at the correct position in the 3D space and hence an optimal point of intersection must be found using stereo triangulation. The optimal stereo triangulation procedure is explained in the following section.

The homogeneous scale factor in Equation 3 is eliminated first by a cross product resulting in the linear vector equations:

$$\begin{aligned} p_L \times Q_L \cdot P &= 0, \\ p_R \times Q_R \cdot P &= 0. \end{aligned} \tag{4}$$

The above two equations can be combined into a homogenous linear equation in P as:

$$AP = 0 \tag{5}$$

Equation 5 defines P only up to an indeterminant scale factor, whereas we seek a non-zero solution for P. We assume that the solution for the 3D point P exists at coordinates (x, y, z, 1) in the 3D space and not at infinity. Based on this assumption, we can reduce the homogenous Equation 5 into a set of six non homogenous equations with three unknowns. A least-squares solution can be computed using *Singular Value Decomposition* (SVD), which is a unit singular vector corresponding to the smallest value of P in the three Cartesian axes x, y and z [13]. The method is computationally efficient and can achieve subpixel accuracy of 3D object reconstruction.

B. Object Orientation Estimation

The 3D points of the four book corners computed using optimal stereo triangulation are denoted by P_i, i = 1,2,3,4. The first book corner is assumed to be the top-left one, while the other corners are ordered in a clockwise manner, as stated in Section 3. The 3D points are reconstructed with respect to the World coordinate system located at the base of the manipulator arm. The object reference frame O corresponds to the intersection of the diagonals of the object, represented by the line vectors $\overrightarrow{P_1P_3}$ with $\overrightarrow{P_2P_4}$, as illustrated in Figure 5. The 3D position of the object reference frame is used. Since books are considered to be planar objects, their rotation can be estimated by computing the orientation of the plane formed by any of the three book corner points. The x axis of O is set parallel to vector $\overrightarrow{P_3P_2}$, whereas the y axis is parallel to vector $\overrightarrow{P_3P_4}$. The z axis is calculated as the cross product of vectors between the x and y axes, as shown in Figure 5. The unit vectors for the obtained principal axis x, y, and z are then computed in Cartesian coordinates. The rotation matrix can be thought of sequence of three rotations, one about each principal axis. The orientation of an object is given by the rotation matrix *Rot* formed using the scalar component of each unit vector representing the x, y and z axes.

$$Rot = \begin{bmatrix} w_{11} & w_{12} & w_{13} \\ w_{21} & w_{22} & w_{23} \\ w_{31} & w_{32} & w_{33} \end{bmatrix}, \tag{6}$$

where, each column represents the unit vectors along x, y and z, respectively.

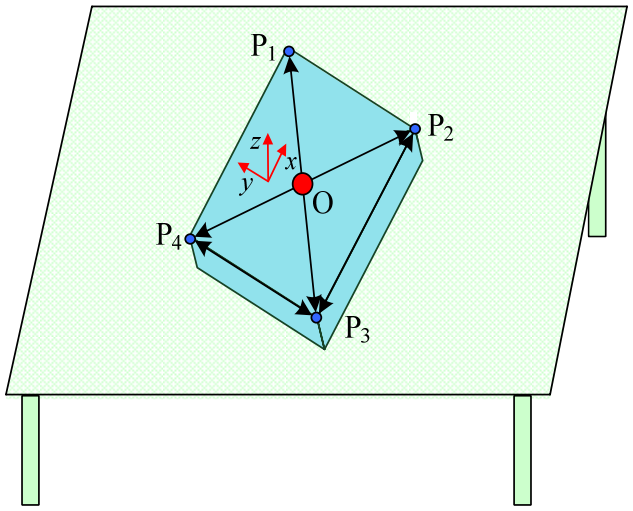


Figure 5: 3D virtual representation of a book object on the library desk and its 3D feature points.

The orientation of the object with respect to the World coordinate system is given by the corresponding Euler angles [13]:

$$\begin{cases} \Phi = \arctan(w_{31}/w_{32}), \\ \Theta = \arccos(w_{33}), \\ \Psi = -\arctan(w_{13}/w_{23}). \end{cases} \tag{7}$$

where Φ, Θ and Ψ give the object's orientation along the *x*, *y* and *z* axes, respectively.

In the two points based orientation estimation method proposed in [14], the objects are not allowed to rotate simultaneously along multiple axis of the reference coordinate system and object tracking is restricted from 0° till 180°. These shortcomings are overcome by the approach proposed in this paper, since an object plane constructed using three 3D points determines the object's orientation. The main advantage of this method is that the orientation of the object can be tracked from 0 till 360° and the object can be simultaneously rotated around all the three axes.

The object pose is defined by the homogenous transformation matrix which relates the object reference frame O with the World coordinate system. The computed object's reference frame O and its orientation angles are saved in the robot's World Model to be further used by the manipulative algorithms. Since the books are always considered to be placed on the library desk, parallel to the floor, only the orientation Ψ, along the z axis, is taken into consideration by the manipulator.

We use cuboids as geometrical shape primitives to model a book in a virtual 3D space, as it has minimal complexity and sufficient accuracy to represent the actual object model. The size of the book which comprises of length, width and height is calculated from the four reconstructed 3D corner points to model of the object. The location of the library desk is calculated using a marker place on it. This marker can be seen on the bottom-left part of the image from Figure 1(b). Using this information, the actual height of the book is computed without any a-priori knowledge about its size. The computed 3D model is further updated in the World Model and used by the manipulative skills to calculate the optimal object grasping point and also for collision avoidance during path planning tasks [15]. Thus, with the help of the computed 3D information, appropriate object grasping and manipulation can be achieved.

5. Performance Evaluation

The success of object manipulation depends on the accuracy of 3D object reconstruction, which also relies on the precision of 2D feature extraction. Hence, the performance of the ROVIS framework used in the library scenario has been evaluated by comparing the accuracy of 3D reconstruction with the actual ground truth. A test image database comprising of 200 images containing books with arbitrary color, size, and texture were taken at different illumination levels in a library setup. The lighting conditions were varied from 15 to 1200 lx. The objects were imaged both under daylight, as well as under artificial lighting conditions. In the first set of experiments, the position and orientation of the considered books were changed under a constant office lighting level of 500 lx in order to evaluate the accuracy of the proposed 3D reconstruction methods. In the second round experiments, the positions of the books were kept constant under varying illumination conditions in order to estimate the robustness of object recognition with respect to scene uncertainties.

The algorithms described in Section 4 were tested with respect to pose estimation precision. As described in the previous sections, the object reference frame O corresponds to the intersection of the diagonals of the four book corners in the World coordinate system. The 3D reconstruction of the object reference frame O for various book positions on the library desk under a constant lighting level of 500 lx is shown in Figure 6.

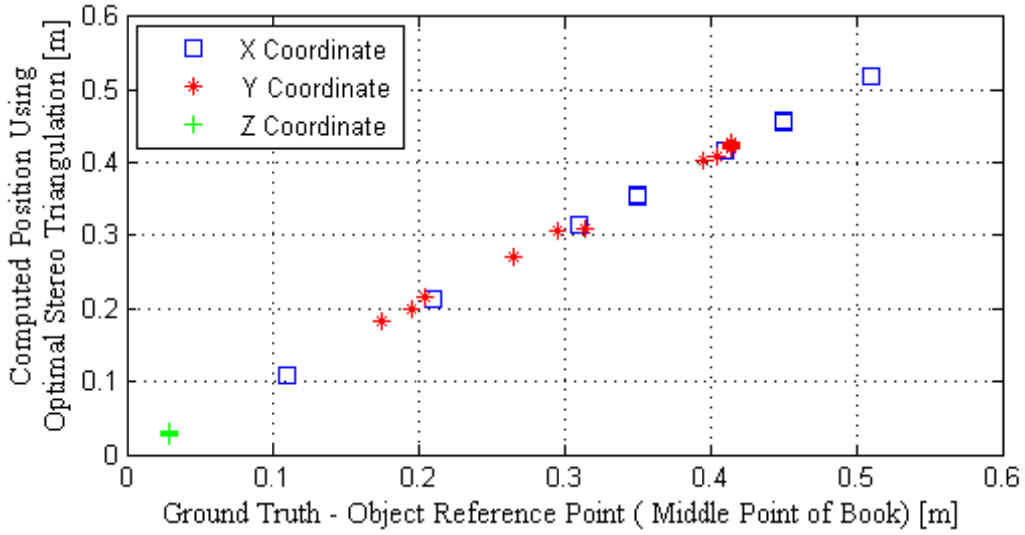


Figure 6: Reconstruction of Object Reference Frame position under constant illumination conditions.

Since book lie on the library desk, we have considered only the object orientation around the z axis (Ψ) of the reference coordinate system. The orientation of the books computed using the proposed plane based object orientation estimation method is illustrated in Figure 7. The presented book rotations are in the range [0°, 360°]. The statistical results of 3D reconstruction are given in Table I. For autonomous vision guided object grasping, the objects should be localized within the desired tolerance limit of 1cm in each direction of the reference coordinate system [14]. From the statistical results we could observe that the error in 3D reconstruction is within the tolerance limits and hence the objects can be dexterously by the manipulator.

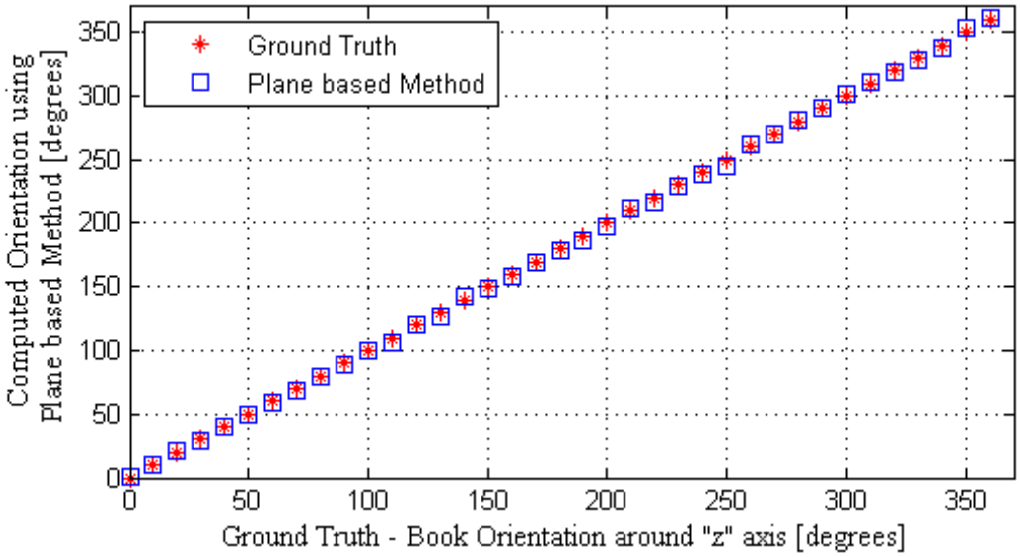


Figure 7: Orientation estimation using the plane based approach.

TABLE I
STATISTICAL RESULTS OF OBJECT 3D RECONSTRUCTION AT VARIOUS OBJECT POSITIONS IMAGED UNDER CONSTANT ILLUMINATION.

	3D Reconstruction Accuracy			
	X_e [m]	Y_e [m]	Z_e [m]	Ψ_e [°]
Max error	0.008	0.011	0.004	3.5
Mean	0.005	0.007	0.001	2
St. deviation	0.002	0.004	0.002	1.5

The position and the orientation error between the computed object reference frame and the actual ground truth for two books using the ROVIS framework and the conventional open-loop object recognition chain under variable illumination conditions are illustrated in Figure 8. The constant segmentation parameters for the open-loop object recognition method were determined under an optimal office lighting level of 500 lx. From the plots and the statistical results given in Table II, we could infer that the ROVIS 3D reconstruction accuracy is high when compared with the traditional open-loop algorithm.

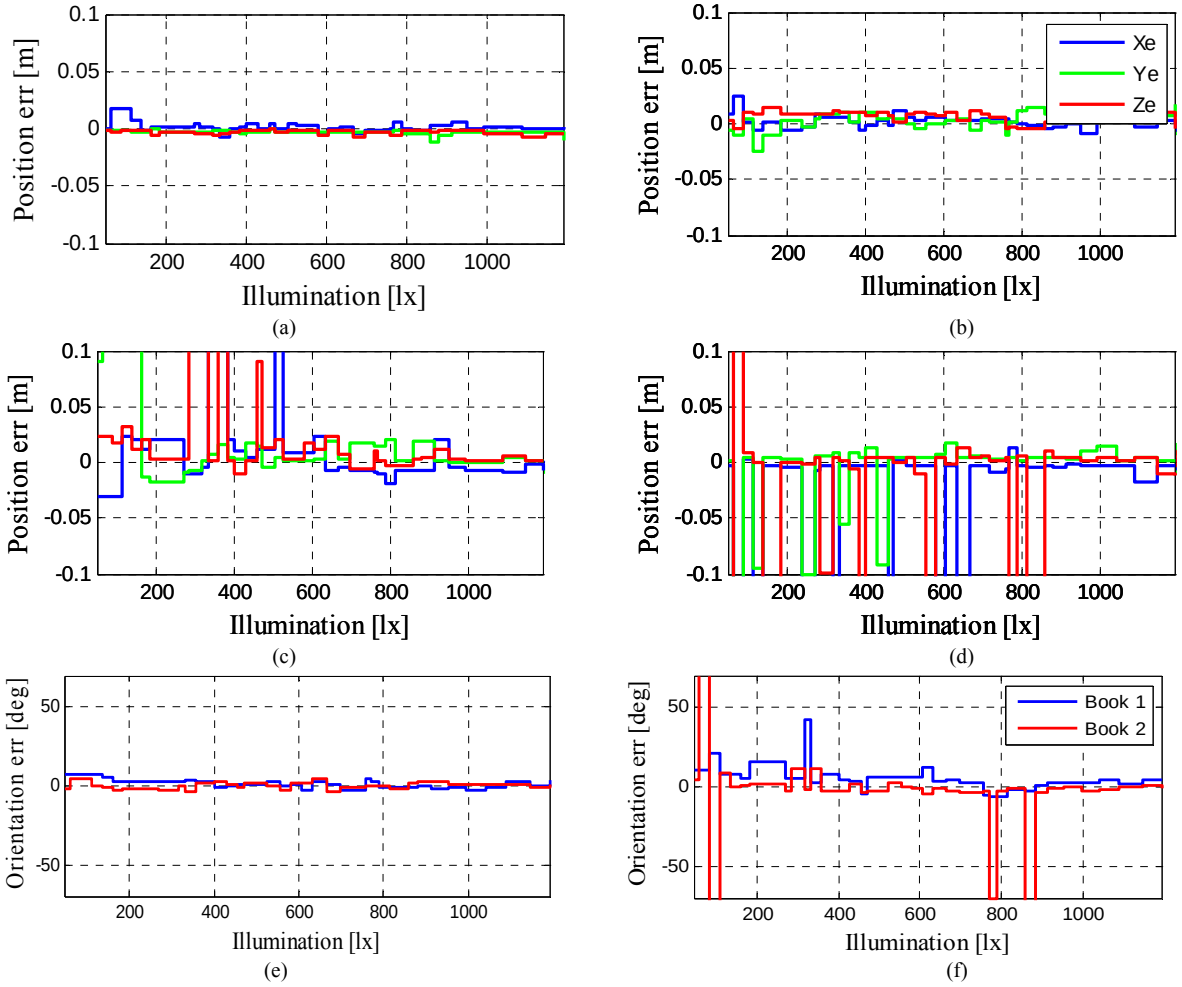
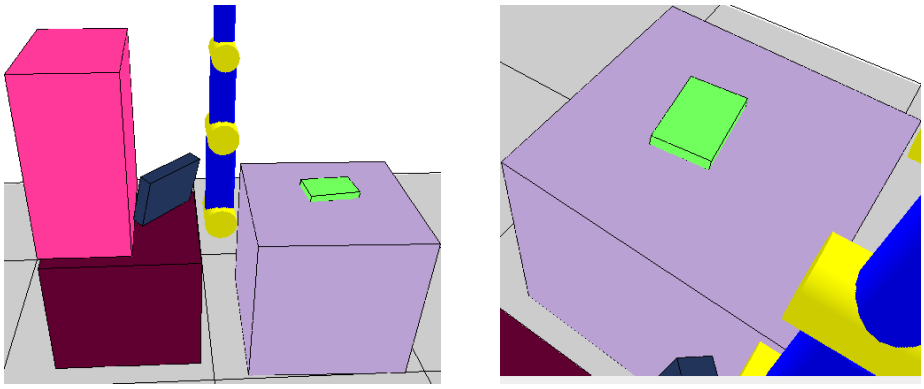


Figure 8: Difference between the real 3D object pose and the calculated 3D object reference pose. ROVIS 3D position error: book 1 (a) and book 2 (b). Open-loop 3D position error: book 1 (c) and book 2 (d). Difference between the real and calculated 3D object orientations: (e) ROVIS. (f) Open-loop.

TABLE II
STATISTICAL RESULTS OF OPEN-LOOP VS. ROVIS BOUNDARY OBJECT RECONSTRUCTION UNDER VARIABLE ILLUMINATION.

	Open-loop 3D Reconstruction				**ROVIS 3D Reconstruction**			
	X_e [m]	Y_e [m]	Z_e [m]	Ψ_e [°]	X_e [m]	Y_e [m]	Z_e [m]	Ψ_e [°]
Max error	0.6294	0.2513	0.7881	34	0.0127	0.0134	0.0069	6
Mean	0.1291	0.1202	0.2493	15	0.0101	0.0054	0.0038	4
St. deviation	0.0360	0.0229	0.0437	8	0.0043	0.0045	0.0011	2

The current spatial positions of the localized objects in the FRIEND' workspace are visualized using a *Mapped Virtual Reality* (MVR) model [16]. Based on the object's size and pose information computed from the 3D reconstruction module, the books are modeled as cuboids and placed within the virtual model reflecting its real world location. The computed 3D model is further used by the robot's manipulation algorithms to compute the optimal object grasping point and also for collision-free motion planning. The MVR model of the localized book on the library desk is shown in Figure 9.



(a) (b)

Figure 9: MVR model of a reconstructed book placed on the library desk: (a) side view (b) top-view.

6. Conclusion and future work

Object recognition and pose estimation are fundamental for autonomous operation of service robotic systems. In this article, a closed-loop control object recognition method has been proposed in order to increase the robustness of object boundary detection. The performance evaluation of the proposed approach was done within the FRIEND library scenario, under varying object positions and illumination conditions. The obtained results demonstrate that the extracted features ensure precise 3D object reconstruction when compared with a conventional open-loop method. The pose estimation algorithm increases the reliability of object localization and uses no a-priori information about the object size. The pose and the geometry of the object are used by a 7-DoF manipulator arm for path planning and object grasping. Future work will include the recognition and 3D reconstruction of books placed in a shelf and also the extension of the orientation estimation algorithm for non-coplanar 3D points.

References

[1] O. Ivlev, C. Martens and A. Graeser, “Rehabilitation robots FRIEND-I and FRIEND-II with the dexterous lightweight manipulator,” *Restoration of Wheeled Mobility in SCI Rehabilitation*, vol. 17, pp.111–123, 2005.

[2] S.M. Grigorescu, D. Ristic-Durrant and A. Graeser, “RObust machine VIsion for Service robotic system FRIEND”, *Proc. of the 2009 IEEE Int. Conf. on Intelligent RObots and Systems*, St. Louis, USA, October, 2009.

[3] A.P. Del Pobil, M. Prats, R. Ramos-Garijo, P.J. Sanz, E. Cervera, "The UJI Librarian Robot: An Autonomous Service Application". *Proc. Of the 2005 IEEE Int. Conf. on Robotics and Automation. Video Proceedings.* Barcelona, Spain. 2005.

[4] F. Viksten, “Comparison of Local Image descriptors”, Proc. of the 2009 IEEE Int. Conf. on Robotics and Automation, Kobe, Japan, May, 2009.

[5] P.A.Beardsley, A.Zisserman, D.W.Murray, “Sequential Updating of Projective and Affine Structure from Motion” , Technical Report, Oxford University, 1998.

[6] R. Hartley, F.Schaffalitzky, “L∞ minimization in geometric reconstruction problems”, IEEE conference on Computer Vision and Pattern Recognition, vol I, pp 769-775, 2004.

[7] D. DeMenthon and L.S. Davis ; “ Model based object pose in 25 lines of code ”, International Journal on Computer Vision, vol.15, pp123-141, 2005.

[8] J.F. Canny ,“A Computational Approach to Edge Detection”, *IEEE Trans. on Pattern Analysis and Machine Learning*, vol. 8, nr. 6, pp. 679-714, 1986.

[9] Majid Mirmehdi, Phil L. Palmer, Josef Kittler, and Homam Dabis. “Feedback control strategies for object recognition”. *IEEE Trans. on Image Processing*, vol. 8, nr. 8, pp. 1084-1101, 1999.

[10] P.V.C Hough, “Methods and Means for Recognizing Complex Patterns”, US Patent 3969654, 1962.

[11] D. Ristic-Durrant, *Feedback Structures in Image Processing*, Shaker-Verlag, 2007.

[12] T. Heyer, S.M. Grigorescu and A. Gräser, “Camera Calibration for Reliable Object Manipulation in Care-Providing Robot FRIEND”, *Proc. of the 2010 Int. Symposium on Robotics,* Munich, Germany, June, 2010.

[13] R. Hartley, A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, 2004.

[14] S.K.Vuppala and A.Gräser, “An Approach for Tracking 3D Object Pose using Two Points ”, 6th International Conference on Computer Vision Systems, Santorini, Greece, 2008.

[15] D. Ojdanic and A. Gräser, “Improving the Trajectory Quality of a 7 DOF Manipulator”, Proc. of the 2008 Int. Symposium on Robotics, Munich, Germany, 2008.

[16] J. Feuser, O.Ivlev, A.Gräser, “ Mapped Virtual Reality for Safe Manipulation in Rehabilitation Robotics”, 2005.