The 2010 IEEE/RSJ International Conference on
Intelligent Robots and Systems
October 18-22, 2010, Taipei, Taiwan

Robust Feature Extraction for 3D Reconstruction of Boundary Segmented Objects in a Robotic Library Scenario

Sorin M. Grigorescu, *Student Member*, *IEEE*, Saravana K. Natarajan, Dennis Mronga and Axel Gräser, *Member*, *IEEE*

***Abstract*—In this paper a vision system for robust feature extraction and 3D reconstruction of boundary segmented objects is presented. The goal of the system is reliable perception of a professional life environment in a scenario of the rehabilitation robot FRIEND. Reconstructed scenes are used to plan object manipulation with a 7-DoF manipulator arm. The robustness of boundary feature extraction is achieved by the means of including feedback control at image segmentation level. The objective of feedback is to adjust the segmentation parameters in order to cope with scene uncertainties, such as variable illumination conditions. Robustly extracted 2D object features are provided as input to the 3D object reconstruction module of the FRIEND vision system. The performance of the proposed approach is evaluated through experiments in the Library scenario of the robotic system FRIEND.**

I. Introduction

A KEY requirement in the field of autonomous service robots is the robust perception of the robot environment aiming at reliable 3D environment reconstruction which is necessary for autonomous object manipulation [1]. Such a service robot, controlled using 3D visual information, is the rehabilitation system FRIEND (*Functional Robot with dexterous arm and user-frIENdly interface for Disable people*), developed at the Institute of Automation, University of Bremen. The goal is to assist disabled and elderly people in their daily and professional life activities (see Fig. 1(a)) [2]. FRIEND is equipped with a 7 Degrees-of-Freedom (DoF) manipulator arm mounted on an electrical wheelchair and a series of sensors used by the robot to understand its environment, like the global stereo camera which views the scene in front of the user of the robot. The robotic system enables the disabled users (e.g. patients which are quadriplegic, have muscle diseases or serious paralysis due to strokes or other diseases with similar consequences for their all day living independence) to perform a large set of tasks in daily and professional life self-determined and without any help from other people like therapists or nursing staff.

(a) (b)

Fig. 1. (a) The rehabilitation robot FRIEND operating in a library. (b) A scene imaged by the FRIEND global camera.

The capabilities of FRIEND have been demonstrated in different scenarios where a large number of consecutive action sequences are performed. These sequences, necessary to fulfil the demands of the robot system's user, are semantically described as robot object handling methods like "pour and serve a drink", "prepare and serve a meal", "fetch and handle a book". In order to plan such actions, reliable visual perception is needed to determine the position and orientation (pose) of the objects in the FRIEND environment. The visual perceptual capabilities of the rehabilitation robot are implemented within the machine vision architecture ROVIS (*RObust machine VIsion for Service robotics*), presented at the IROS 2009 conference [3]. One of the main features of ROVIS is the inclusion of feedback structures at image processing algorithmic level as well as between the vision and other modules of the robotic system. The goal of including feedback structures is to achieve high robustness against external influences, of the individual system units as well as of the system as a whole. The effectiveness of ROVIS has been demonstrated in [3] by the reliable recognition and 3D reconstruction of uniformly coloured objects (e.g. typical household objects like bottles, glasses, meal-trays, etc.) in Activities-of-Daily-Living (ADL) scenarios of the rehabilitation robot FRIEND. The ADL scenarios enable the user to prepare and serve meals or beverages. Within the ADL scenarios, the preparation and serving of meals and drinks by the robot is demonstrated.

Another support scenario developed for the FRIEND robot

Manuscript received August 18, 2010.

S.M. Grigorescu is with the Department of Automation, Transilvania University of Braşov, Mihai Viteazu 5, 500174, Braşov, Romania (phone: +40-268-418-836; email: s.grigorescu@unitbv.ro).

S.K. Natarajan and A. Gräser are with the Institute of Automation, University of Bremen, Otto-Hahn-Allee NW1, 28359 Bremen, Germany (phone: +49-421-218-62470 / 7523; e-mail: {grigorescu, natarajan, ag} @iat.uni-bremen.de).

D. Mronga is with the German Res. Center for Artificial Intelligence (DFKI), University of Bremen, Robert-Hooke-Strasse 5, 28359, Bremen, Germany (phone: +49-421-218-69560; email: dennis.mronga@dfki.de).

takes place in a professional life situation, where the user is working at a library desk equipped with a laser scanner for reading IDs of books and customer IDs. The task of the FRIEND user is to handle outgoing and returned books, as well as other tasks at a library desk. One of the positive aspects of this so-called *library scenario* is the encouragement of the user to interact with other people, thus supporting his recovery and reintegration into professional life. In Fig. 1(a), FRIEND is shown operating in a complex library environment, whereas in Fig. 1(b) the library desk imaged by the stereo camera of the robot can be seen.

In recent years, robotic perception in domestic environments has been treated in a number of publications [4,5]. The number of systems known to examine similar ADL and professional life scenarios as FRIEND is high, so only a few of them will be mentioned here. The UJI librarian robot [6] was designed to detect IDs of books on a shelf. The vision system of this robot considers only the detection of books IDs, followed by their pick up from the shelf using a special designed gripper and hybrid vision/force control. In comparison to [6], the FRIEND Library scenario aims at the recognition and 3D reconstruction of all types of books, placed in cluttered environments, and their grasping using a standard gripper. Learning algorithms for 3D object orientation have been described in [7], whereas in [8,9] databases were used for modelling object grasping. However, the reliability of low-level vision, from which grasping points are calculated, is still an open problem. One approach for recognition of objects with sufficient texture characteristics is the usage of methods exploiting local texture. Such methods, as SIFT model based recognition, were used in [10] to perform full pose estimation of objects. Although SIFT like algorithms perform well on objects with high amount of texture, uniform coloured ones are difficult to detect. Recently, active range sensing technologies, such as laser scanners [11] and time-of-flight cameras [12] have been used in understanding complex domestic settings. In [11] a method for scene segmentation and reconstruction from 3D point clouds provided by a laser scanner has been described. However, such methods are strictly dependent on the range data quality provided by the sensing device. Sensed depth information can have different error values depending on the sensed surface.

In this paper, a contribution to robust 2D recognition and 3D reconstruction of objects in complex scenes is given. The main idea of the approach is to use classical image processing techniques enhanced by including feedback control at low-level image processing, an idea also tackled previously in the computer vision community [13,14]. In contrast to [13,14], this paper approaches the inclusion of feedback control in image processing from the point of view of robotic manipulation, where precise 3D reconstruction is crucial for optimal object grasping.

The paper is organized as follows. In Section II the complete ROVIS object recognition and 3D reconstruction chain for books detection is described, followed in Section III by its closed-loop extension. In Section IV performance evaluation is given through experimental results. Finally, conclusions are discussed in Section V.

II. The ROVIS Machine Vision Architecture

From the image processing point of view, the objects of interest in the FRIEND scenarios have been classified into two categories: *Container Objects*, which represent relatively large objects in the scene (e.g. library desk, refrigerators, microwave ovens, etc.) and *Objects to be Manipulated* (e.g. books, meal-trays, bottles glasses, etc). In this paper recognition of containers is taken for granted, while the focus is on recognition and 3D reconstruction of textured objects, such as books. Since the objects of interest are textured, their recognition is based on boundary based segmentation.

Fig. 2 shows the block diagram of the ROVIS machine vision architecture [3]. The vision system is composed of three main modules: hardware, initialization and the object recognition and reconstruction chain.

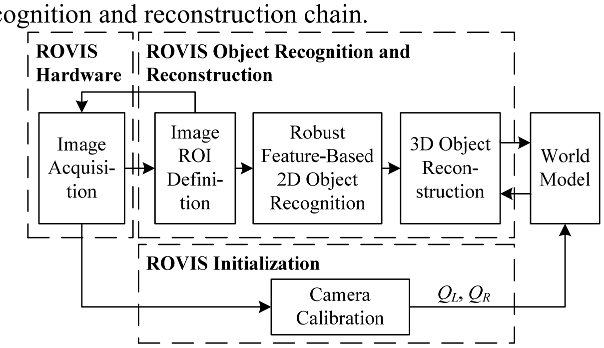


Fig. 2. Block diagram of ROVIS.

The ROVIS hardware is composed of a Bumblebee® global stereo camera attached to a 2-DoF pan-tilt head unit, mounted on a rack behind the user, as illustrated in Fig. 1(a). The system is initialized through the ROVIS Camera Calibration procedure [15], which calculates 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 lenses of the stereo camera, respectively. In the rest of the paper, the reference coordinate system will be named as the *World* coordinates. As it will be explained in Section III, the projection matrices are used by the 3D Object Reconstruction module to calculate the pose of the object to be manipulated with respect to the world coordinates. The calculated calibration data is further stored in the so-called *World Model*, which acts as a database for storing runtime visual information.

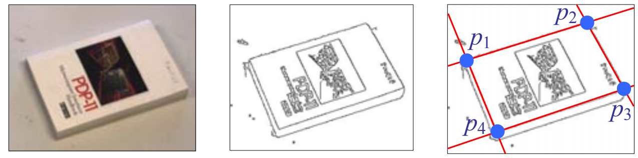
The first step in the ROVIS Object Recognition and Reconstruction chain is the definition of the image *Region of Interest* (ROI) containing the objects to be recognized. In case of the FRIEND Library scenario, the image ROI is defined by detecting a marker placed on the left side of the library desk [15]. If the marker is not visible in the field-of-view of the camera, the vision system changes the orientation of the camera by adjusting the angles of the pan-tilt head. In this way ROVIS visually searches for the marker, as indicated by the feedback arrow connecting the Image ROI Definition with the Image Acquisition module. The image ROI is set by knowing the geometry of the library desk and the position of the marker on it. The final ROI, that bounds the desk, will contain one or more book objects that have to be reconstructed in a 3D space. If more books are recognized, the user of FRIEND has to select the desired one through a special human-machine interface device [3]. On the calculated image ROI, 2D feature-based object recognition is applied with the goal of extracting so-called *object feature points*. These points are inputs to the 3D reconstruction module which calculates the object's pose. Finally, the 3D reconstructed objects are stored in the World Model and used by the manipulative algorithms to plan manipulation tasks [16]. The success of object manipulation depends on the precision of 3D object reconstruction, which, on the other hand, relies on the precision of 2D feature extraction.

III. Object Recognition and 3D Reconstruction Based on Boundary Segmentation

The recognition of textured objects, like books, is mainly done through methods that detect their boundaries, or edges [17]. Such a method, named edge segmentation, aims at classifying as foreground object pixels the ones that lie on the edges of objects. Pixels are considered as edges if they lie on sharp local changes in the intensity of an image. The output of segmentation is a binary image where foreground object pixels have the value 1 (black) and background pixels the value 0 (white), as seen in Fig. 3(b). One main drawback of pure, raw, edge segmentation is that often breaks between edge pixels are encountered. For the case of line edges, a way around this problem is to evaluate the colinearity of binary edge pixels. This evaluation can be performed using the so-called *Hough transform* [18] which converts the raw edge pixel data to a parameter space suitable for collinearity analysis. In order to distinguish between raw edge lines and lines calculated with the Hough transform, the latter will be referred to as *Hough lines*. In Fig. 3(c), the red lines represent the extracted Hough lines.

Boundary based 3D object reconstruction, using edge segmentation and Hough transform, is implemented in ROVIS as the image processing chain shown in Fig. 4. The goal of the processing chain is to parallel process left and right stereo images in order to extract the feature points of the objects to be manipulated and used them to reconstruct their pose. In case of books, feature points are represented by the four book corners:

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



(a) (b) (c)

Fig. 3. Boundary feature point extraction. (a) Input image of a book. (b) Edge segmented image. (c) Hough lines and feature point extraction.

where p_{Li} and p_{Ri} represent book corners in 2D image coordinates (x_i, y_i) in left and right images, respectively. The blue circles in Fig. 3(c) represent the extracted feature points. As convention, the first book corner is assumed to be the top-left one. Using the calculated 2D feature points from Eq. (1) and linear stereo triangulation [19], the position of the book in 3D space can be reconstructed and saved in the World Model as the homogenous 3D object coordinate:

$$\mathrm{O} = (x, y, z, 1)\,, \tag{2}$$

where x, y and z represent the 3D object's position O along the three axes of the Cartesian space. O is calculated as the intersection of the four corners of a book, as it will be explained in Section III.D.

A. Boundary Segmentation

In ROVIS, boundary segmentation is considered as the combination of raw edge segmentation and Hough transform. Edge segmentation is performed using the *Canny edge detector* [20]. Basically, the Canny algorithm is performed through two main steps: filtering of the input intensity image with the derivative of Gaussian of a scale σ and thresholding the filtered image by the so-called *hysteresis thresholding*. The binary edge detected image is calculated with the help of low T_L and high T_H thresholds, aiming at detecting strong and weak edges, where the weak edges are included in the output image only if they are connected to strong edges. The low threshold can be expressed as a function of the high threshold as:

$$T_L = 0.4 \cdot T_H\,. \tag{3}$$

An example of a Canny edge segmented image can be seen in Fig. 3(b).

One drawback of using only raw edge detection for boundary object extraction is that very often the obtained contour edges are not connected, that is, they have small breaks between the edge pixels. This phenomenon happens due to noise in the input image, non-uniform illumination and other effects that introduce discontinuities in the intensity image [17]. The Hough transform [18] is a method used in linking edge pixels based on shape. Although any shape can be expressed by the so-called generalized Hough transform, in practice, because of computational expenses, shapes like lines or ellipses are used. In this paper, the goal is to extract the lines that bound a book. These lines are calculated by estimating the collinearity of raw edge pixels. The Hough transform maps the binary edge pixels to the

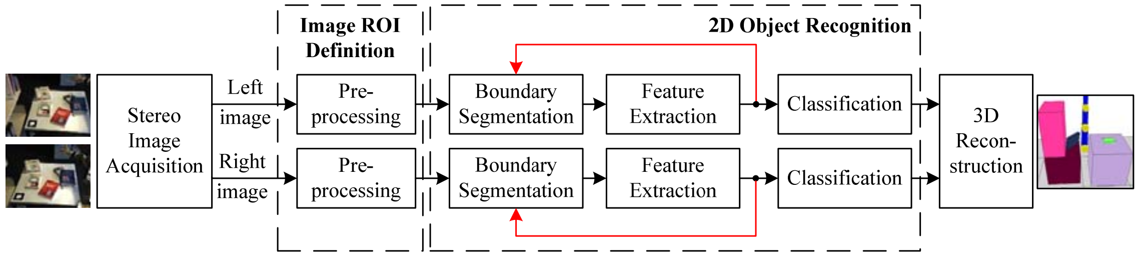


Fig. 4. Boundary segmented object recognition and 3D reconstruction chain in ROVIS.

so-called *accumulator cells*. Initially, the accumulator cells are set to 0. For every foreground pixels that lies on a line, a specific cell of the accumulator is increased. The higher the number of pixels that lie on a line, the higher the values of the corresponding accumulator cell is [18]. Since the value of the accumulator entries reflects the collinearity for all foreground edge pixels, it is meaningful to threshold it, so to consider as Hough lines only the ones which have an accumulator cell value higher than a specific threshold T_{HG}. In Fig. 3(c), the red lines represent the detected Hough lines.

The values of Canny and Hough transform thresholds are usually used as constant values which poses problems in variable illumination conditions. In Section IV, a closed-loop method for automatic adjustment of these thresholds, as illumination during image acquisition changes, is introduced.

B. Boundary Feature Extraction

In order to establish the transformation from object segmentation to feature extraction, the extracted Hough lines have to be grouped into structures forming candidate objects, also called *reference models*. A book reference model is represented as a combination of parallel and perpendicular lines, forming the standard shape of a book. The construction of the candidate solutions represent a combinatorial expansion of the angles between detected lines in the input image ROI. The angle of a line, ν, is measured with respect to the x axis of the 2D image plane. Ideally, between two parallel lines the difference in their angles should be 0, whereas for perpendicular lines $\pi/2$. Considering the camera's viewing angle and possible image distortions, the decision of classifying two lines as parallel or perpendicular has been done by introducing two offsets. Two lines are considered to be parallel if the difference in their angles with respect to the x axis is smaller then 12°, or $0.209\,\text{rad}$. Likewise, two lines are considered perpendicular if the angles difference varies in the small interval $[\pi/2-0.209, \pi/2+0.209]$. Further, if one line pair is perpendicular to another one it is considered as a *candidate object solution* and added to the solutions vector $N_\#$.

C. Classification

Because of image noise and texture, not all the candidate solutions in vector $N_\#$ represent real objects, which are books. The purpose of the classification procedure described here is to distinguish between spurious candidate solutions, called *negatives*, and real objects, named *positives*. This has been achieved with the help of a Minimum Distance Classifier [29] and different extracted object features, as follows:

- Relative object area $A_r = \dfrac{A_{obj}}{A_f}$,
- Eccentricity (object's width-to-height ratio) R_{wh},
- Relative number of object pixels $R_{px} = \dfrac{N_f}{N_b}$,

where A_{obj} represents the object area bounded with extracted lines and A_f the area of the whole image ROI. N_f and N_b correspond to the number of foreground and background pixels covered by the Hough lines in the binary segmented image, respectively. The above mentioned features were combined in two Euclidean distance measures forming the two positive and negative object classes:

$$D_{pos} = \sqrt{(A_r - a_1)^2 + (R_{wh} - b_1)^2 + (R_{px} - c_1)^2}\,, \tag{4}$$

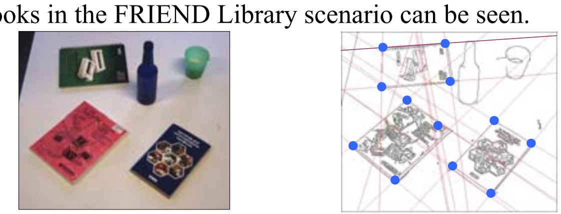
$$D_{neg} = \sqrt{(A_r - a_2)^2 + (R_{wh} - b_2)^2 + (R_{px} - c_2)^2}\,, \tag{5}$$

where the coefficients in Eq. (4-5) have the values:

$$\begin{aligned} &a_1 = 0.03; \quad b_1 = 0.9; \quad c_1 = 1.1; \\ &a_2 = 0.01; \quad b_2 = 0.7; \quad c_2 = 2.5. \end{aligned}, \tag{6}$$

The coefficients in Eq. (6) have been determined heuristically using a number of 50 positive training samples and 50 negative ones. An object is considered positive and classified as a book if $D_{pos} > D_{neg}$.

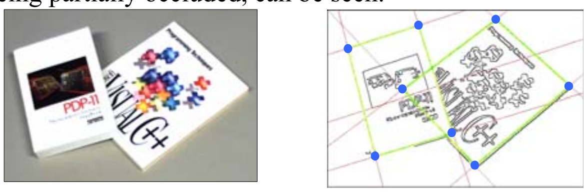
The intersections of the Hough lines of a detected object give the four object feature points p_{Li} in a 2D image. In Fig. 5 the extraction of object feature points from recognized books in the FRIEND Library scenario can be seen.



(a) (b)

Fig. 5. Extraction of object feature points. (a) Input image. (b) Recognized objects and their feature points.

In real world environments, object grasping and manipulation based on visual information has to cope with object occlusion. One advantage of the Hough transform is that it can cope with partial occlusion [18], that is, candidate solutions are found although objects overlap each other by a certain degree. In ROVIS, a partially occluded object is considered a positive if it is occluded by less than 30% of its own area. In Fig. 6 the result of recognizing two books, one being partially occluded, can be seen.



(a) (b)

Fig. 6. Object recognition with partial occlusion. (a) Input image. (b) Recognized objects and their feature points.

D. 3D Object Reconstruction and Manipulation

Once the feature points of the objects to be manipulated have been extracted in both left and right stereo images, the 3D reconstruction module is used to calculate the pose of the object.

The first step of reconstruction is to calculate the 3D position of the four extracted feature points using stereo triangulation [19]. The determined 3D points are then used to compute the orientation of the object, as it will later be explained. In the following, the 3D reconstruction of one feature point P will be detailed. The calculation of the other three points is made using the same procedure. We consider a linear mapping from 3D points to 2D image coordinates in both left and right stereo images. The relationship between the considered 3D point P and its perspective 2D image projections from Eq. (1) is given as:

$$\begin{aligned} p_L &= Q_L \cdot \mathrm{P}, \\ p_R &= Q_R \cdot \mathrm{P}, \end{aligned} \tag{7}$$

where p_L and p_R represent 2D image feature point coordinates in the left and right stereo images, respectively. Q_L and Q_R are the left and right camera projection matrices determined during ROVIS initialization at the Camera Calibration stage, as seen in Fig. 2. Q_L and Q_R are defined as the product of the intrinsic and the extrinsic camera parameters.

Using the projection matrices the homogeneous scale factor from Eq. (2) is eliminated by a cross product resulting in the linear vector equations:

$$\begin{aligned} p_L \times Q_L \cdot P &= 0, \\ p_R \times Q_R \cdot P &= 0. \end{aligned} \tag{8}$$

The least-squares solution for Eq. (8) is a unit singular vector corresponding to the smallest value of P in the three Cartesian axes x, y and z. This value is obtained using Singular Value Decomposition (SVD) [19].

The object's reference frame O, which gives the object's 3D position, is calculated with respect to the World coordinate system. As seen in Fig. 7, O corresponds to the intersection of the diagonals of the object, that is of the line vectors $\overrightarrow{P_1P_3}$ with $\overrightarrow{P_2P_4}$. The x axis of O is set parallel to vector $\overrightarrow{P_2P_3}$, whereas the y axis is parallel to vector $\overrightarrow{P_1P_2}$. The z axis is calculated as the cross product between the x and y axes.

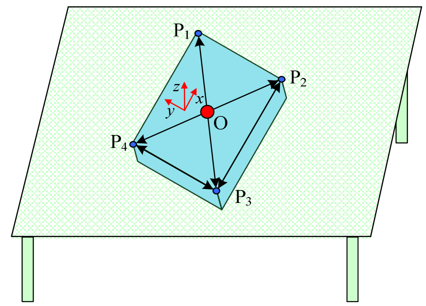


Fig. 7. 3D virtual representation of a book object on the library desk and its 3D feature points.

The object's 3D position O is used by the manipulation algorithms to calculate the object grasping point. For grasping purposes, additional information regarding the object's orientation is needed. The orientation of an object is given by the Φ, Θ and Ψ Euler angles, determined using the orientation of the unit vectors of $\overrightarrow{P_3P_2}$ and $\overrightarrow{P_3P_4}$:

$$(\Phi, \Theta, \Psi) = f(\overrightarrow{P_3P_2}, \overrightarrow{P_3P_4}), \tag{9}$$

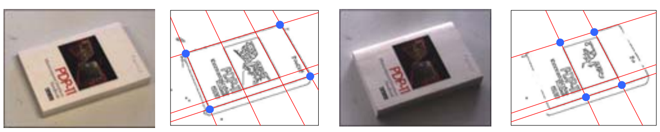
where Φ, Θ and Ψ are the object's orientation along the x, y and z axes, respectively [19].

Both the object's reference frame O and its orientation angles are saved in the robotic system World Model to be further used by the manipulative algorithms. Since a book is always placed on the library desk, parallel to the floor, only the orientation Ψ, along the z axis, is taken into consideration by the manipulator. With the help of the computed 3D information, appropriate object grasping and manipulation can be achieved. The grasping of a book from a library desk is performed as follows: first the manipulator slides the detected book on the edge of the library desk, thus setting a small part of the book away from the desk. The book is grasped by this small part using the manipulator's gripper. An example movie showing this procedure accompanies this paper.

IV. Feedback Control Extension of Boundary Segmentation

As shown in Section III, a crucial requirement for reliable object manipulation using visual information is the robust extraction of object feature points used for 3D reconstruction. This requirement is strictly related to the quality of boundary segmentation. A boundary segmented image is said to be of good quality if the calculated object boundaries lie on its real boundaries. The extension of the boundary segmentation algorithm presented in Section III employs the idea of inclusion of feedback structures at image processing level to control the quality of the segmented image ROI. The idea behind this approach is to change the parameters of image ROI segmentation in a closed-loop manner so that the current segmented image ROI is driven to the one of reference quality independently of external influences.

One still open problem in computer vision and particularly in robot vision is the robustness of object recognition with respect to variable illumination conditions. This problem is exemplified in Fig. 8, where object feature points are extracted using the segmentation method presented in the previous section. The method uses constant parameters determined in reference artificial illumination conditions. As can be seen from Fig. 8, the feature points are reliably extracted for the case of artificial illumination. When the same constant segmentation parameters are used for recognizing the object in changed illumination (e.g. daylight) the output result is incorrect. In order to cope with the influence of illumination, a closed-loop boundary segmentation method is proposed.



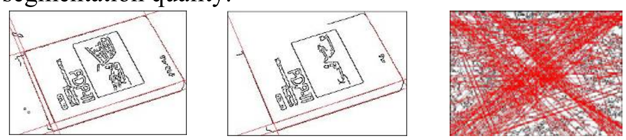
(a) (c) (b) (d)

Fig, 8. 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.

The concept of closed-loop image processing, also treated in [13,14], has been successfully applied by the authors for controlling region based segmentation in variable illumination conditions [3], with the purpose of recognizing uniformly colored household objects for visual guided object grasping. In this section, a closed-loop control method for boundary segmentation is proposed. In such a feedback control system, the control signal, or actuator variable, is a parameter of image segmentation and the controlled variable a measure of feature extraction quality. This concept is shown in Fig. 4 by the feedback connecting feature extraction with boundary segmentation. In the following, the choice of the actuator and controlled variables for the proposed closed-loop system is detailed, along with the design of the feedback controller. In Section V of this paper it is demonstrated that the proposed closed-loop boundary segmentation method contributes directly to the robustness of the 3D reconstruction algorithm presented in Section III.

A. Choice of the actuator and controlled variables

In closed-loop image processing, actuator variables are those parameters that directly influence the image processing result [21]. Since the boundary segmentation method used in ROVIS is composed of the Canny edge detector and the Hough transform, the actuators are chosen as the parameters that most strongly influence these operations. The result of Canny edge detection is dependent on the choice of low T_L and high T_H thresholds. For the sake of clarity, T_L is considered to be a function of T_H, as shown in Eq. (3). For the rest of this paper, the Canny thresholds will be referred only to T_H. On the other hand, the Hough transform is strongly influenced by the value of the accumulator threshold T_{HG}. Different feature point extraction results for different values of the Canny and Hough transform thresholds are shown in Fig. 9. As can be seen, only a proper choice of these thresholds gives a result of good boundary segmentation quality.



(a) (b) (c)

Fig. 9. Feature points extraction results for different values of T_H and T_{HG}. (a) Optimal segmentation. (b) Under-segmentation. (c) Over-segmentation.

In order to control the T_H and T_{HG} thresholds a measure of boundary segmentation quality had to be defined. This quality measure is to be used as the controlled variable in the proposed closed-loop system. As said before, a good boundary segmented image ROI is one where the detected Hough lines lie on the real object's edges. Such a quality measure can be calculated at the feature extraction level, where detected Hough lines are combined in order to form the objects candidate solutions vector $N_\#$. The equation of the proposed measure is:

$$y = \begin{cases} e^{N/N_{\max}} \cdot \sum_{n=0}^{N_\#} \frac{N_f(n)}{A_{ROI}}, & \text{if } N \leq N_{\max} \\ 0, & \text{if } N > N_{\max} \end{cases} \tag{10}$$

where N represents the total number of Hough lines, $N_\#$ the number of candidate solutions and $N_f(n)$ the number of foreground pixels covered by the Hough lines of the n^{th} object, normalized with the area of the image ROI, A_{ROI}. Having in mind the computational burden of the Hough transform [17], the maximum number of lines allowed in an image ROI is set to a constant value $N_{\max}$. The exponential term in Eq. (10) is introduced in order to force feature extraction with a minimum amount of Hough lines. Hence, y decays to zero when the number of Hough lines increases. For the feature extraction results in Fig. 9, the corresponding values of y are given in Table I.

As can be seen from Table I, the higher the value of the quality measure y is, the better the image ROI segmentation

TABLE I.
VALUES OF THE PROPOSED QUALITY MEASURE FOR THE FEATURE EXTRACTION RESULTS FROM FIG. 9.

	Optimal segmentation	Under-segmentation	Over-segmentation
y	0.5728	0.4821	0.2934

quality is. To investigate the system's input-output controllability when considering the thresholds T_H and T_{HG} as the actuator variables and the measure y as controlled variable, boundary segmentation was applied on the image from Fig. 5(a). The value of T_H was varied in the interval $[0,255]$, whereas the value of T_{HG} in the interval $[0,100]$. For each combination of thresholds $[T_H, T_{HG}]$ the controlled variable y was measured. The input-output result can be seen in Fig. 10. Optimal feature extraction corresponds to the combination of thresholds which maximize the variable y. It can be observed from Fig. 10 that different combinations of thresholds yield the same value of the quality measure. This is because the same optimal feature extraction result can be achieved with different values of T_H and T_{HG}.

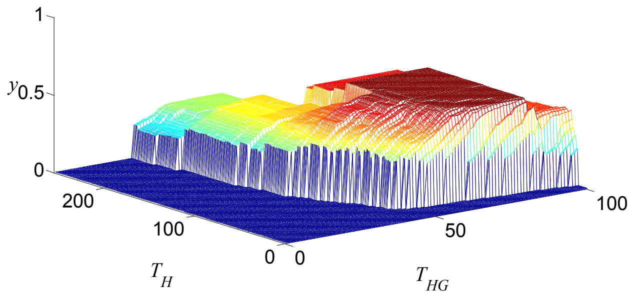


Fig. 10. The quality measure y vs. Canny T_H and Hough transform T_{HG} threshold.

B. Feedback Control of Boundary Feature Extraction

In Fig. 11 the block diagram of the proposed closed-loop boundary segmentation method is displayed. In the presented system, the reference value of the chosen controlled variable is not explicitly known, since the goal is to develop a method able to detect objects independent of their sizes, color, or texture information. The objective of the control structure from Fig. 11 is to find the maximum value of the controlled variable y. This is achieved through a feedback optimization process using an appropriate extremum seeking algorithm. Since, as said before, optimal feature extraction is achieved for different thresholds values, the extremum seeking algorithm stops when the gradient of the surface in Fig. 11 reaches 0. Optimal determined values of the Canny and Hough transform thresholds ensure that a reliable input is given to the feature extraction module which directly influences the precision of 3D object reconstruction.

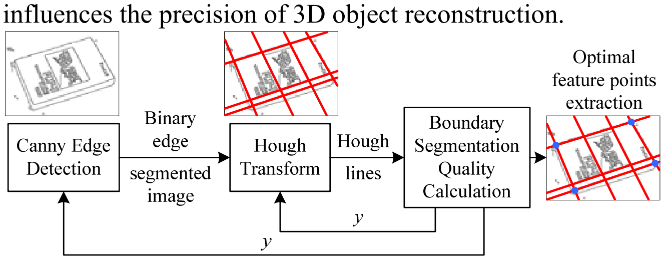


Fig. 11. Cascade control structure for robust boundary segmentation.

V. PERFORMANCE EVALUATION

The objective of the ROVIS system is robust 3D object reconstruction for appropriate object grasping and manipulation [16]. Therefore, the evaluation of ROVIS in the Library scenario of the system FRIEND was performed with respect to the precision of final 3D reconstruction under variable illumination conditions. A number of 30 book objects arranged in different configurations (e.g. with orwithout partial occlusion, placed among other types of objects, etc.) were used. 200 test images were acquired in the illumination interval $[15lx, 1200lx]$. This range of illumination corresponds to a variation of the light intensity from a dark room lighted with candles (15lx) to the lighting level of an office and above. According to the European law UNI EN 12464 the optimal lighting level of an office has a value of 500lx. Each captured image was segmented using the open-loop boundary detection method described in Section III.A and with the improved closed-loop algorithm from Section IV. The constant segmentation parameters of the open-loop method were determined at the artificial illumination value of 500lx. The obtained features were used for further 3D reconstruction, that is, the calculation of the detected objects reference frames and orientations. These 3D results were compared to the manually determined pose of the objects. In Fig. 12(a,b,d,e), the 3D position errors for two book objects are given. It can be seen that 3D errors obtained using ROVIS are smaller than the ones of the open-loop algorithm. As explained in Section III.D, since the book lies on a library desk parallel to the ground, only the evaluation of the orientation along the z object axis was taken into account, namely the value of Ψ from Eq. (9). The variation of the orientation error, Ψ_e, can be seen in Fig. 12(c,f). The statistical measures of achieved errors in all the experiments are given in Table II.

VI. CONCLUSION

In this paper, a 2D recognition and 3D reconstruction image processing chain for robust feature points extraction of boundary segmented objects has been proposed. The goal of the method is to provide a robust input to the 3D reconstruction module. Obtained 3D object poses are used by a 7-DoF manipulator arm for path planning and object grasping. The benefit of inclusion of feedback control at feature extraction level for improvement of boundary segmentation has been demonstrated through experiments with the rehabilitation robotic system FRIEND. The proposed method increased the reliability of the robotic system and its usage in professional life by disabled patients, under variable illumination conditions.

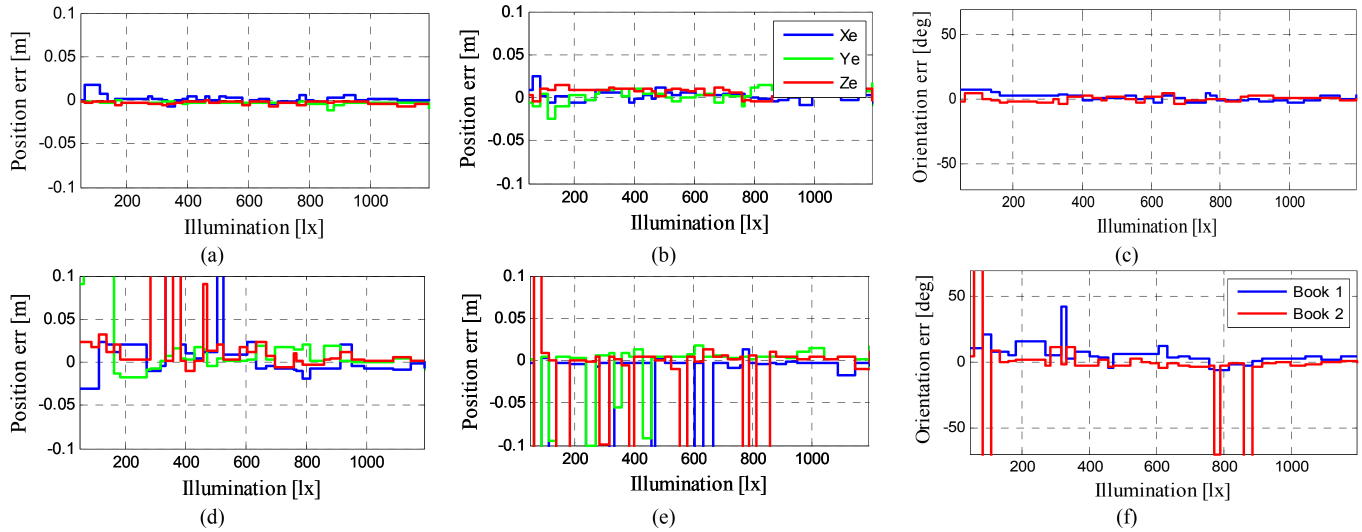


Fig. 12. Difference between the real 3D object position and orientation and the calculated 3D object reference frame. ROVIS 3D position error: book 1 (a) and book 2 (b). Open-loop 3D position error: book 1 (d) and book 2 (e). (c) ROVIS 3D orientation error. (f) Open-loop orientation error.

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

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

REFERENCES

[1] D. Kragic and H.I. Christensen, "Advances in robot vision", *Robotics and Autonomous Systems*, 52, pp. 1-3, 2005.

[2] 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.

[3] 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.

[4] T. Asfour, P. Azad, N. Vahrenkamp, K. Regenstein, A. Bierbaum, K. Welke, J. Schröder and R. Dillmann, "Toward Humanoid Manipulation in Human-Centred Environments", *Robotics and Autonomous Systems*, pp. 54-65, 2008.

[5] D. Kragic, M. Björkman, H. Christensen and J.-O. Eklundh, "Vision for Robotic Object Manipulation in Domestic Settings", *Robotics and Autonomous Systems*, 52, pp. 85-100, 2005.

[6] 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.

[7] A. Saxena, J. Driemeyer and A.Y. Ng, "Learning 3-D Object Orientation from Images", *Proc. of the 2009 IEEE Int. Conf. on Robotics and Automation*, Kobe, Japan, May, 2009.

[8] U. Klank, M. Zeeshan Zia and M. Beetz, "3D Model Selection from an Internet Database for Robotic Vision", *Proc. of the 2009 IEEE Int. Conf. on Robotics and Automation,* Kobe, Japan, May, 2009.

[9] C. Goldfeder, M. Ciocarlie, H. Dang and P.K. Allen, "The Columbia Grasp Database", *Proc. of the 2009 IEEE Int. Conf. on Robotics and Automation,* Kobe, Japan, May, 2009.

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

[11] R.B. Rusu, N. Blodow, Z.C. Marton, Michael Beetz, "Close-range Scene Segmentation and Reconstruction of 3D Point Cloud Maps for Mobile Manipulation in Domestic Environments", *Proc. of the 2009 Int. Conf. on Intelligent RObots and Systems*, St. Louis, USA, October, 2009.

[12] J. Kuehnle, A. Verle, Z. Xue, "Integration of 6D Object Localization and Obstacle Detection for Collision Free Robotic Manipulation", *Proc. of the 2009 IEEE Int. Conf. on Advanced Robotics*, Munich, Germany, June, 2009.

[13] M. Mirmehdi, P.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.

[14] J. Peng and B. Bahnu, Closed-loop Object Recognition Using Reinforcement Learning", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 20, nr. 2, pp.139-154, 1998.

[15] 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* (to be published)*,* Munich, Germany, June, 2010.

[16] 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.

[17] R. Gonzalez and R. Woods, *Digital Image Processing*, 3rd edition, 2007.

[18] P.V.C Hough, "Methods and Means for Recognizing Complex Patterns", US Patent 3969654, 1962.

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

[20] 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.

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