Reliable Hand Camera based Book Detection and Manipulation in Library Scenario

Stefan Heyer, Christos Fragkopoulos, Torsten Heyer, Axel Gräser
Institute of Automation, University of Bremen, Germany
Email: {sheyer, cfragkopoulos, theyer, ag}@iat.uni-bremen.de

Abstract—In this paper a new approach for detection, grasping and placing of books is described. The location as well as book’s size, color and textures are unknown. Therefore the autonomous book detection and grasping is a big challenge. In this work, the proposed approach is implemented in the scope of a library scenario in which a service robotic system operates. The proposed approach uses stereo vision for coarse approach to the next book, which should be grasped, and a hand camera (mounted on the gripper) based positioning strategy to achieve a high success rate of grasping. The full concept was implemented and tested on the service robot FRIEND. It should support a handicapped person to work and to perform all manipulative tasks. These contain the autonomous book detection on a book cart, the grasping and placing of the book on a book holder and later onto the cart again. The benefit of our approach is documented based on experimental results.

I. INTRODUCTION

The robotic system FRIEND, displayed in Figure 1 [1]–[4], was developed at the Institute of Automation (IAT) at the University of Bremen (Germany) in order to support disabled and elderly people in their activities of daily as well as in professional life. It consists of a wheelchair platform equipped with a 7-DoF robot arm and a series of sensors to capture information from the surrounding environment, a Bumblebee® stereo camera, a 3D IR-camera and an eye-in-hand camera mounted on the gripper of the manipulator. Since service-robotic systems generally work in unknown and clustered environments all object which should be grasped or manipulated have to be acquired first by the sensors.

In the latest research project ReIntegraRob [5] a person with quadriplegia will be enabled to return to profession by means of FRIEND. The person should be able to perform the tasks of a librarian without being dependent on a personal assistance. FRIEND will fulfill all manipulative tasks since this cannot be done by the person itself. In this project a handicapped person supported by robot assistance will work in the same manner as a healthy person. The task which is done by the disabled person is to catalog stocks of books from the archive and to insert them into the on-line data base of the library. The cataloging procedure itself is done by the disabled person and cannot be automated, whereas all manipulative tasks are done by FRIEND. For several reasons described later it turns out that the task is more difficult then it seem on first impression, especially when a failure rate below 1% is requested. To reduce the task complexity the whole scenario is simplified as much as possible to achieve a high success rate.

The simplifications are: a) The books are arranged in a special way on the book cart, alternating portrait and landscape placement and book front to the robot, so that the books can be grasped easily by the gripper and can be placed in the right orientation on the book holder. b) A horizontal black stripe is used as a marker on the vertical board of the book cart to support positioning of the gripper in front of the right most book. Book cart and book holder are detected by ARToolKit markers [6] which are placed on the objects (see Figure 1), allowing a simple and reliable 3D positioning. c) The pages of the books will be flipped using an automated device, which is under development (in Figure 1 only a provisional solution is shown).

The books are supplied on a book cart which has to be detected by FRIEND’s vision system before a coarse approach of the robot arm to the book cart is done. Then a fine approach using the hand camera is performed in order to find the right most book on the first shelf. After grasping the book it is placed on an automated book holder, detected by vision, which is able to open or close the book. The pages of the book will be flipped using the already mentioned automated device. The user can control this automated book holder directly through the FRIEND system in order to find the right pages with the information necessary for the cataloguing. After cataloguing the book is taken from the book holder, placed again on the book cart on the second shelf and the next book is chosen.

The paper is organized as follows. After a state of the art overview we describe our book detection algorithm in
Section III. In Section IV it is shown how the task is performed from manipulative point of view. In Section V the benefit of our approach is evaluated based on test results. Finally, conclusions and outlook are presented in Section VI.

II. STATE OF THE ART

To the best of our knowledge there exist up to now only few systems for book detection [7]–[9]. Those systems use easy identifiable markers or codes on each book or optical character recognition for book localization on a book shelf in a library. In [10] each book is equipped with a RFID tag and the book size is stored previously in a database. After detection of the tag all book information are accessible and used for book grasping. In [11] a similar approach is presented, which uses bar codes on the books instead of the RFID tags. In the work from Tomizawa [12]–[14] only simulation experiments have been presented. Compared to other approaches, the work presented here does not depend on any identifier label or priory knowledge and is able to detect and grasp arbitrary books of different color and size. Especially books with the same color and size are often delivered from magazine for retro catalogization. This means, that old books will be added to the on-line database of the library. In difference to the vision-guided grasping strategy by Sanz [15] our approach uses the vision system only for book detection and an end-effector mounted force-torque sensor for collision detection with the environment. During grasping and placing of the books such a contact between gripper and obstacles is partly desired in our approach (see Section IV).

III. BOOK DETECTION VIA HAND CAMERA

The stereo camera is approximatively between 1.2 and 2 meters away from the book cart and is therefore not able to provide precise position information of the right most book. So a hand camera mounted on the top of the gripper is used to get an image from a closer distance to the desired target object. During searching for the right most book the gripper follows a trajectory in a constant distance (around 5cm) in front of the book cart (in Figure 2 from point $A$ via $B$ to point $PB$). The distance of the hand camera to the right most book ($Z$-direction) is not constant due to the alternating portrait and landscape placement of the books and the different book dimensions. The detection of the right most book via hand camera consists of two algorithms. The first one, called Black Stripe detection with Hand Camera (BSwH), is used to compute the intersection point $PS$ (robot world, 3D) and its 2D equivalent $BS$ in the image plane between the black stripe and the right most book on the shelf, respectively. The second one, called Book Detection with Hand Camera (BwH), determines the upper right book corner $R$ (image plane, 2D) and the slope of the book. In Figure 3 an overview of the presented book detection algorithm is given. $M$ denotes the center point of the image (2D) and $PS$ the detected end position of the black stripe calculated in the previous iteration (see Figure 2). More details are given in the following subsections.

A. Coarse approach to book cart and to right most book

After book cart detection using stereo vision the robot arm is moved to a fixed position at the beginning of the book cart (position $A$ in Figure 2). Via image processing techniques like threshold segmentation, Canny edge detection and line detection using Hough algorithm the end-effector follows (see Section IV) the black stripe until the right most book is visible in the hand camera image (position $B$ in Figure 2). The detected end position of the black stripe $BS$ (2D) is mapped...
to its 3D world position PS using the known tool center point of the gripper B and the position of the black stripe on the book cart.

B. Pre-processing

There are many influences which disturb the quality of the book detection like image noise or varying illumination conditions. The books can be quite different from each other, generally not uniformly colored and have maybe text fields. Also miscellaneous, where all books are equal due to size, color and texture, can occur. In order to decrease these influences image pre-processing is performed. An overview of the main steps of the image pre-processing procedure is shown in Figure 4. This pre-processing phase delivers two results, which will be independently processed by the book detection algorithm (see following subsection).

Gray-scale images are used since experiments have shown that they are sufficient for book detection. Smoothing and dilation operations are applied to eliminate image noise and small contours coming from structures on the books. The processed image is converted to an edge image using Canny’s algorithm. Starting from this point the image processing is continued in two different chains independently from each other in order to achieve a higher book detection reliability. In the first branch almost vertical lines, which represents book borders or book pages are extracted via Hough line detection. The following line connection procedure is shown in Algorithm 1.

Coping with varying illumination conditions is one of the biggest challenge in all fields of image processing and still an open field for research. In the second example in Figure 5 (lower row) the branch with the Hough Line detection yields a worse result than the other approach, since there the book dimensions in the lower part are not detected correctly. In this case these affects not the further process, since the upper right book corner is used as reference point.

C. Book detection algorithm

In Algorithm 2 the main steps of the book detection algorithm are shown. For the two different input images generated by the pre-processing phase (see Figure 4) the same book detection algorithm is used. At the end the results of both branches are compared and the right most book is selected. The right most book is at the moment the best possible choice. Rarely the book cover itself is detected as the most right book candidate. This happens, if the book is turned a bit on the book shelf. This influences the detection of the reference point R only marginal and leads mostly to a successful grasping of the book. If both branches deliver two different outputs, i.e. the difference is bigger than the minimal book thickness, experimental results have shown, that these were caused primarily by the fact that one branch has rejected the most right book and detects instead of that the neighbouring book
The detected upper right book corner resp. the upper right book point $R$, in case if the upper book border is not visible inside the image, is determined. Based on the deviation between the detected book point $R$ and the image middle point $M$ (see Figure 6) a correction movement for the gripper is calculated and applied. The used contour detection algorithm is very sensitive against small gaps, i.e. it detects no contours, which outlines have gaps. For this reason filling of line gaps by the pre-processing phase is needed. The main steps of the book detection algorithm consist of eliminating contour defects, adding proper contours to a book candidate list, rejecting book candidates, which are not the desired ones, and selecting of the right most book candidate (see Algorithm 2).

Algorithm 2 Book detection algorithm

1: Abbreviations:
2: $m_1 =$ Minimal size of a book.
3: $m_2 =$ Maximal size of a book.
4: for all Input images do
5: Detect contours
6: for all Contours $C$ do
7: Eliminate contour defects
8: if $m_1 < \text{Size}(C) < m_2$ then
9: Adding $C$ to list of book candidates
10: end if
11: end for
12: end for
13: for all Book candidates $B$ do
14: Analyze $B$ w.r.t. book criterias
15: if $B$ is not a book then
16: Eliminate $B$
17: end if
18: end for
19: Select right most candidate

IV. MANIPULATIVE TASK EXECUTION

The automatic placement of the book on surfaces like book cart and book holder implies a clustered structure. The motion planning algorithm, used in this work, is called CellBiRRT [16]. Any motion described in this work is done using that approach.

For the rest of the paper the following abbreviations are going to be used: $\{W\}$ refers to world coordinate system which is attached to the base of the robot arm, $\{H\}$ represents the book holder’s, $\{B\}$ the book’s, $\{G\}$ the end effector’s and $\{C\}$ the book cart’s coordinate system. The end effector’s $\{G\}$ frame is calculated with the help of forward kinematics from the current robot arm configuration. The $\{H\}$ and $\{C\}$ result from the acquisition with image processing. The translation matrix $T_{CG}^{HW}$ describes the location (position and orientation) of the book w.r.t. the world coordinate system. For all images, the green arrows denote the positive $X$ axes, the blue arrows the positive $Y$ axes and the red ones the positive $Z$ axes.
A. Grasping a book from book cart

In order to grasp a book on the book cart the end effector should be first placed in front of the book cart. The $BSwH$ starts from this point. The end effector frame is calculated by the formula:

$$T_G^W = T_G^W \cdot T_B^G$$

(1)

As it can be seen from Figure 7(a) the $X_C$ axes of $\{C\}$ is parallel to the $Z_C$ axes of $\{G\}$. Also the $Z_C$ axes is parallel to the $X_G$ axes. Therefore the rotation part of the $T_G^W$ can be calculated as follows: rotation $90^\circ$ around $Y_C$ axes and rotation $180^\circ$ around the new $X_C$ axes: $T_G^C = \text{Rot}_Y(90) \cdot \text{Rot}_X(180) \cdot T_{\text{trans}}$. The last part contains a translation in $\{G\}$ frame. The result is presented in Figure 7(a). If a book is successfully grasped the actual $T_{G_{\text{actual}}}^C$ is stored in the database for later usage. The latter accelerates the total motion, since the detection of the black stripe is not necessary for the next trial.

The book, when it is grasped, and the book cart hold the same orientation. When a book is grasped, the relative frame $T_B^W$ is calculated using the formula: $T_B^W = (T_G^C)^{-1} \cdot T_B^C$. This relative frame is used later in the sub scenario of placing the book on the book holder. The Figure 7(b) illustrates an example and the coordinate systems.

B. Place book on book holder

Figure 8(a) illustrate an example of placing a book on the book holder. The coordinate systems are: $T_G^W$, $T_B^W$, and $T_B^H$ corresponding to the $\{H\}$, $\{G\}$ and $\{B\}$ coordinate systems. The strategy for placement has the following steps:

- **Learning (offline) phase**: The end effector is placed manually to the desired position. The location of the book is calculated from the formula: $T_B^W = T_G^W \cdot T_B^G$. The desired relative frame between the book and the book holder $T_H^W$ is calculated from the equation: $T_B^W = (T_H^W)^{-1} \cdot T_B^W$. This procedure is done only one time during the development.

- **Execution (online) phase**: During the execution phase, the final end effector location is calculated using the following formula:

$$T_G^W = T_H^W \cdot T_B^H \cdot T_B^G$$

C. Place book on book cart

The strategy for placing a book on the book cart is the same as for the first part of grasping a book from the book cart. The end effector moves parallel to the cart (see Figure 8(b)) and the orientation is calculated similar to the equation 1. Afterwards a motion parallel to the $Y_C$ axes is done (yellow arrow) until the mounted force-torque sensor detects a force. Then the book is placed down and the gripper is opened.

V. EXPERIMENTAL RESULTS

The results of two experiments are shown in Figure 9 and Figure 10, which illustrate the achieved position accuracy of book detection via hand camera and its reliability. The right upper book corner $R$ is used as a reference point. The single view of the hand camera is not able to provide depth information, i.e. with the hand camera only a correction along $X-$ and $Y$-direction in the gripper coordinate system can be done (see Figure 2). Due to missing depth information no tolerance measures for positioning of the gripper in millimeters can be given. The only way is to define, how much deviation [in pixel] between $R$ and $M$ (see Figure 3 and Figure 6) is allowed to achieve successful book grasping on the one hand side and on the other side convergence of the positioning process. Empirical experiments have shown, that 10 pixels tolerance in $X-$ and $Y$-direction for the book detection algorithm is suitable for our application and it corresponds to 6-21mm depending on the recent book and its location. Portrait or landscape leads to a varying distance between hand camera and book between 15-25mm (determined by measuring the minimal and the maximal distance between hand camera and books placed on the book cart). Smaller tolerance values lead to instabilities of the positioning process, especially for books, which are more far away, i.e. the gripper positioning procedure starts oscillating around the final position. In all trials of both test series the book detection algorithm reaches the tolerance area, i.e. the detected right upper book corner $R$ was always inside the tolerance area around the image middle point $M$ and book grasping can be started. In Figure 9 and Figure 10 the deviation between detected upper right book corner $R$ and the real book corner are displayed. The yellow bar indicates the given tolerance area.

Figure 9 shows the achieved reliability for positioning accuracies for 20 consecutive trials on the same book. The positioning in $Y$-direction is very stable and always within...
the tolerance. The tests show that here the tolerance interval can be reduced in order to position the gripper more accurate in front of the right most book. In X-direction only 85% of the trials are inside the tolerance area. Noise which results from shadows on the books or the background (trial 2, 7, 19) affects the quality of the results and leads to wrong detection of the upper book corner.

In a second experiment (Figure 10) the book detection algorithm was tested on 40 different books, which were placed on the book cart alternating in portrait or landscape format (starting with portrait). Also here the positioning in Y-direction is very stable and all except one are inside the tolerance area (trial 30). In trial 30 the book cover of a thick book was wrongly detected as an additional book and this causes the big deviation of 23mm in Y-direction. In X-direction 15 (37.5%) of the trials are outside the tolerance area. The algorithm yields in two cases inaccuracies. The first case is when the upper rear border line of the book is detected instead of the frontal one (see Figure 11(a)-(b)). The second case causes in the fact that for the most books the book case is a bit bigger than the book pages (up to 7mm in X-direction). But the presented algorithm can detect only the upper end of the book pages and as consequence there could be an inaccuracy. This second case is shown in Figure 11(c)-(d).

Noises resulted from shadows on the books or the background affects the quality of the results and leads to wrong
detection of the upper book corner. Neighboring landscape books (trial 1, 9) can cast a cloud over the current portrait book and can cause, that this shadow is detected as the upper book boundary. As consequence the upper book border is detected too low. Whereas portrait books (trial 16, 24, 36) can cast a cloud over the background of the book cart. Especially if the upper book corner does not stand out much from the background, this shadow can be detected wrongly as the upper book boundary. As consequence the upper book border is detected too high. If the deviation is too much this leads in the first case to a failure during grasping, since the gripper is above the book. In the second case the book can be grasped, but not in the best manner at the upper corne. It is grasped deeper as the examples in Figure 12 show. In the second experiments eight books could not be grasped either due to inaccurate detection of the book or since the gap to the next book was too small for grasping so that the gripper hits the next book. The latter comes from the fact, that the size of the books is unknown. As consequence it is not precisely known, how much the gripper must be opened in order not to contact the neighboring book. In future the neighboring book should be detected by stereo vision and these information will be combined with the information received by the hand camera to get a higher reliability of book detection and to control the opening of the gripper for grasping.

Fig. 12. Different books were grasped: (a)-(c) correct grasping and (d) wrong grasping (not at the corner)

VI. CONCLUSION

In this paper a new approach for reliable book detection and grasping by a service robotic system in a library scenario is shown. For coarse approach of the robot arm stereo vision is used and the fine approach and grasping procedure is supported by an eye-in-hand camera. The evaluation showed that the robotic system FRIEND was able to grasp different books of unknown size, color and shape with a high rate of success. Only when the illumination conditions are bad or very thin books should be grasped, our approach fails. Experimental results showed that books with thickness less than 1.5cm cannot be grasped.

To get rid of these lacks the presented approach will be improved in future and disparity information will be combined with the information received by the hand camera. Also filter algorithms for shadow removal will be integrated in order to increase the robustness of the proposed book detection method also under bad illumination conditions. A detection of the neighboring books would support the grasping procedure and helps to avoid collisions with the next book during grasping. Algorithms for distance estimation between hand camera and book will be investigated in order to set the tolerance parameters for the positioning based on this, so that fixed parameters can be avoided. This will increase the robustness and reliability of the proposed algorithm.

REFERENCES