

An Autonomous Assistant Robot For Book Manipulation in a Library^{*}

Ramos-Garijo R, Prats M, Sanz PJ, Del Pobil AP

Computer Science Department

Jaume-I University

Castellón, Spain

al024080, al019137, sanzpj, pobil@uji.es

Abstract - *This paper represents work in progress towards a complete system working to assist users in a library. With this aim, the system must be capable to looking for a specific book in a shelf, asked by any user, and whether it is found, deliver it as soon as possible to the user. To get its objectives the system integrates automatic object recognition, visually guided grasping, and force feedback, among other advanced capabilities. Implementation details about the main modules developed presently are shown. Finally, after success in preliminary results obtained in our Campus, we are encouraged to follow working in this way to obtain the complete prototype.*

Keywords: Autonomous Manipulation; Pattern Recognition Robots; Visually-Guided Grasping.

1 Introduction

Presently, a very active research on service robots has been detected. This can be observed looking at the programs followed by the most important conferences around over the world in robotics (ICRA, IROS, etc.). Commonly, underlying to this kind of systems a mobile robot is needed. Many applications until now have been reported, focused on tour-guide robots, cleaning robots and so on. But is more unusual the application in a real life scenario, like a library, where manipulation is a must. In this last case a mobile robot arm is necessary to manipulate the books on a shelf. In summary, we can distinguish three main components in such a systems: a mobile platform, including navigation capabilities; a robot arm, suitable for autonomous manipulation tasks; and a user interface, letting a very high level of interaction with the system.

1.1 State of the art

Focus on these aforementioned systems, some recent progress has been achieved. In particular, and with the aim to reduce the many difficult tasks included in this kind of systems, a teleoperated solution was presented by Tomizawa and coworkers [4], from the University of Tsukuba, in Japan. Although the final objective in our

work is different to that proposed by Tomizawa et al., some strategies can be useful to our project. Note that we are not interested in a posterior digitization of printed materials, as it is proposed by Tomozawa et al., but all the previous strategies necessary to guide the robot towards the bookshelf, identify and manipulate any book, etc., it is the same for us. On the other hand, some works has been developed in an autonomous manner, such as that from the Johns Hopkins University, untitled "Comprehensive Access to Printed Materials" (CAPM). Its main objective, presently under progress, will be allow for real-time browsing of printed materials through a web interface. So, with this aim, an autonomous mobile robotic library system has been developed to retrieve items from bookshelves and carry them to scanning stations located in an off-site shelving facility. Nevertheless, until now only simulation experiments have been presented by their authors Suthakorn et al. [3]. It is important to clarify the main difference between [3] and [4], from a robotics point of view. Although the two systems are using internet for the user interaction, only in [4] the user is within the system control loop, namely it is a teleoperated system.

1.2 Motivation and goals

In this paper, an autonomous solution for the robotics librarian is proposed. The aim will be to retrieve a book required by any user, and bringing that book to the user, whether it was founded in the corresponding bookshelves. Until now, same previous pieces, related to this problem, have been developed in our lab, such as: robot navigation strategies; user interfaces based on voice commands; or visually guided grasping modules [2], among others. In particular, this paper is focused on the computer vision and grasping modules, necessary to achieve the objective proposed.

The rest of the paper has been organized in the subsequent sections. The overall system description is presented in section 2. The main modules developed until now, namely vision and grasping, are shown in sections 3 and 4, respectively. Section 5 reports some preliminary results. And finally, some concluding remarks are presented in section 6.

^{*} 0-7803-7952-7/03/\$17.00 © 2003 IEEE.

2 Overall System Description

Taking into account the objective to achieve, in the following, see Fig 1, the system functionality, focus on the manipulation strategy, is summarized. We can distinguish in this figure the following items:

1. “User Input”. That is responsible of letting the user a very high level of interaction with the system. In particular, voice commands are permitted (e.g. “bring me the Introduction to Robotics by McKerrow”). Before to pass to the next module, the translation from the initial user commands to the corresponding label is achieved, by using a database.
2. “Scanning Mechanism”. This item includes several modules necessities to looking for the required book. In particular, we must process via computer vision, any label found in all the books located on the different bookshelves. An Optical Character Recognition (OCR) module, based on [1], is included here.
3. “Success?”. Each one that a label is identified, it is compared with that required by the user in his input. While that comparison do not be successful, the “scanning mechanism” continues. But, whether the comparison is O.K., the system executes the “grasping module”.
4. “Grasping module”. This is the responsible of the autonomous manipulation of the book finally found. To achieve its objectives needs processing force and visual sensory information.

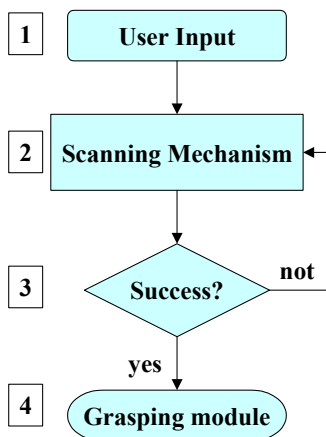


Figure 1. The flow chart showing the system functionality.

In the following, the low level details, about all the modules described here, will be shown.

3 The Computer Vision Module

With the aim to achieve a natural manner for users interacting with the robot, a voice recognition module is operating jointly with the computer vision system. The complete flow chart can be seen in Fig 2. Where, on the left, the input voice commands associated to a book required by the user is shown. The next step is looking for the code associated to this book in the corresponding data base. Bearing in mind, that this codes are physically present in all the book’s labels. In Fig 2-right, the parallel image process can be appreciated. The step (1), it consists in the capture of two images of different resolution (i.e. high, 1280×1032; and low 320×40) at the same time. The step (2) computes the location (i.e. centroid, etc.) of all the labels found in the image. Moreover, the vertical borders between books are detected. The objective of the step (3) is to establish a suitable region of interest over the image with high resolution. Each region of interest is very close to each physical label found in the image. After a process of magnification, step (4), each label is processed by the OCR module, step (5), that obtain the final code associated with the book. Thus, if the code so found coincides with the input code by the user, the suitable information is passed to the grasping module. The most important steps aforementioned are: the location module of labels, and the OCR Module.

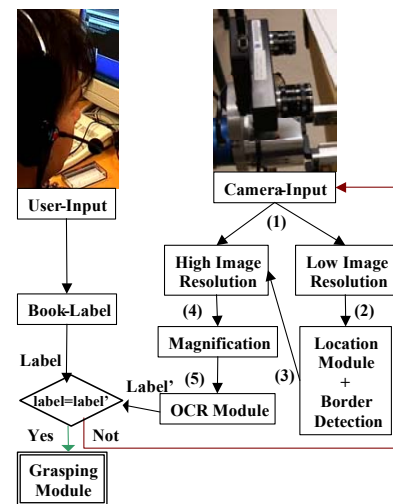


Figure 2. The “flow chart” to locate and identify the label of a book, before grasping it.

3.1 Location Module

The main steps of this algorithm are the following:

- 1 Looking for the interest region within the image. (see Fig 3).



Figure 3. The “interest region” capture.

- 2 Border detection of books found in the image. (see Fig 4).



Figure 4. Border detection of books.

- 3 For each blob image associated to a book the following must be computed: (see Fig 5-a)

- 3.1 Histogram Equalization (see Fig 5-b).
- 3.2 Binarization by means of thresholding techniques (see Fig 5-c).
- 3.3 Remove short areas from the image resulted in the last binarization process (see Fig 5-d).

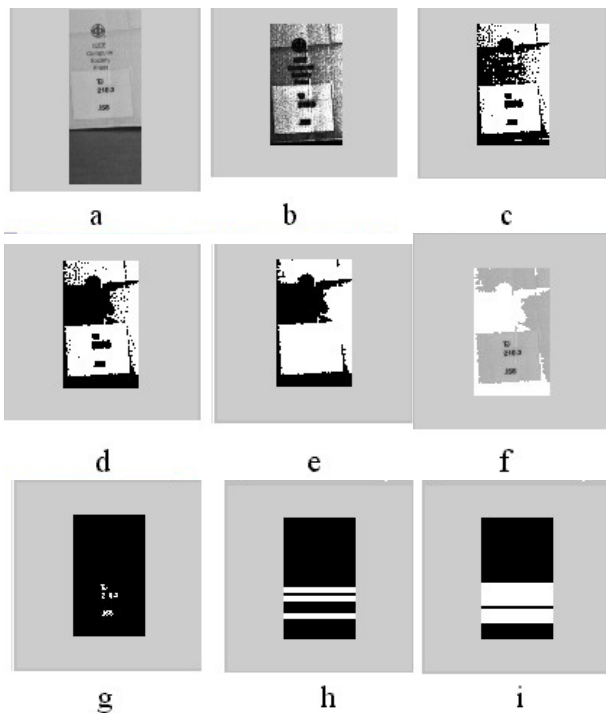


Figure 5. The different results associated to the step 3.

- 3.4 Fill the possible holes within the last binary image obtained (see Fig 5-e).
- 3.5 By means of the binary image adjust the pixels to 255 (i.e. corresponding to the white color), those that previously have zero values in the binary image (see Fig 5-f).
- 3.6 Binarization by means of thresholding techniques over the modified image, looking for black areas in the image (see Fig 5-g).
- 3.7 Process all the pixels of the binary image, and each one that a pixel with value 1 is found, adjust to 1 all the pixels of the row where this pixel was found (see Fig 5-h).
- 3.8 Dilate the binary image. (see Fig 5-i)

For instance, a result obtained following the aforementioned procedure can be seen in Fig 6.

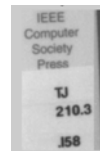


Figure 6. The final result associated to the step 3.

3.2 OCR Module

This module is the responsible to translate the pixels associated to images of possible labels, to authentic text labels, formed by characters, as can be found in the corresponding data base. Thus, the following steps are necessary to achieve this objective:

1. Read as input the number of fields and the fields associated to the label of the book that is looking for.
2. For each label passed through the label detection module, including the number of the book associated to this label, do:
 - 2.1. Trimming the label area in the high resolution image (see Fig 7-a).
 - 2.2. Reduce, as much as possible, the size of the label, without losing of including text. To achieve this objective the following steps are necessary:
 - 2.2.1. Detection of contours by means of Sobel masks (see Fig 7-b).

2.2.2. Process all the pixels of the binary image, and for each pixel found with value 1, adjust all the pixels belonging to its column to 1 (see Fig 7-c).

2.2.3. Dilate the binary image obtained in the previous step (see Fig 7-d).

2.2.4. Remove from the label those areas that have black color in the binary image (see Fig 7-e).

2.2.5. Edges detection with Sobel masks over the image of label obtained previously (see Fig 7-f).

2.2.6. Process all the pixels from the image obtained in 2.2.5., and for each pixel with value 1, adjust all the pixels belonging to its row to value 1 (see Fig 7-g).

2.2.7. Dilate the binary image obtained (see Fig 7-h).

2.2.8. Remove from the label those areas that have black color in the binary image (see Fig 7-i).

2.3. Magnify the label obtained in the step 2.2.8., twice its original size, by means of a digital zoom based on a bi-cubic interpolation (see Fig 7-j).

2.4. Generate 20 binary images of the image obtained in 2.3., using 20 different thresholds belonging to the following sequence of values: 50, 55, 60, 65, ..., 150 (see Fig 8).

2.5. Process, by using the OCR module, all the images obtained in 2.4.

2.6. Whether each field belonging to the label of the book required, a corresponding text with identical characters is found in the binary images, it means that the book has been found. In this case the information about the edges of that book will be passed to the grasping module. In other case, the next label will be processed.

Related to the information passed to the grasping module, once the label is detected, it is noticeable that a pair of grasping points in opposite sides of the book are defined. These grasping points define a line, the grasping line, so the orientation of the gripper, and location of the two fingers are fixed in the work area of the robot arm.

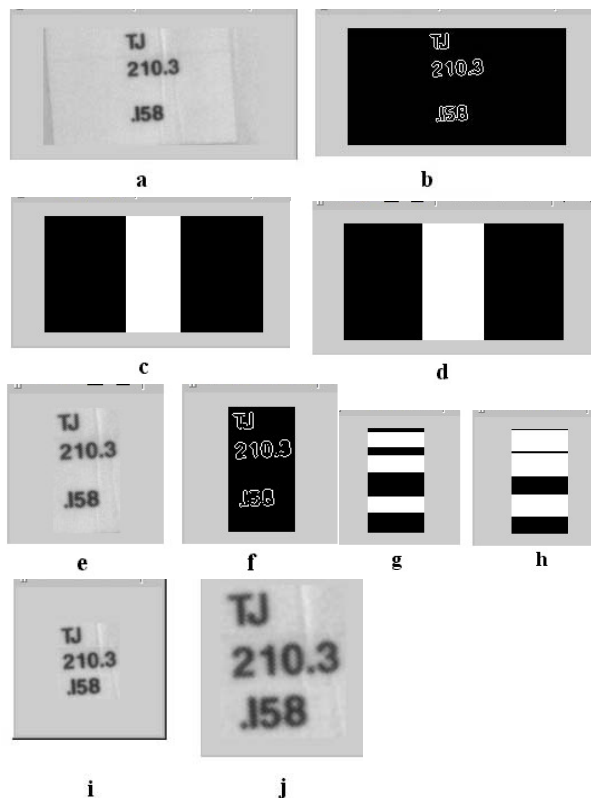


Figure 7 The different stages to prepare the suitable input for the OCR module.

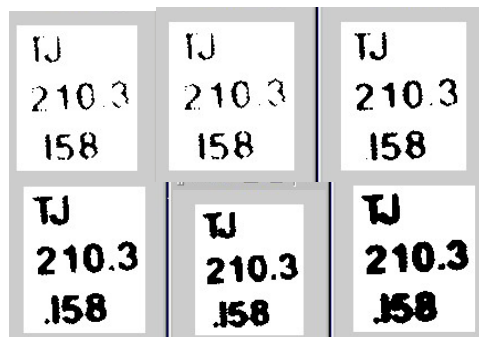


Figure 8. A set of images, extracted from the sampling process, used as input in the OCR module.

4 The Grasping Module

An intensive work has been necessary to adapt our robot system to the requirements involved on extraction of books from shelves. As can be seen in Fig 9, special fingers have been designed to accomplish this task. Moreover, presently some initial hypothesis must be assumed:

- ✓ Books must be graspable by our robot arm. So, its physical properties (i.e. length, wideness, weight, etc.) will be always suitable to the available manipulation capabilities (i.e. geometry of the gripper, etc.).

- ✓ Books are located in a vertical position, approximately, over the shelf. Nevertheless, a little inclination degree is also allowed.
- ✓ The side of the books from the robot point of view, oriented towards outside the shelf, are all of them, approximately, in the same spatial plane.
- ✓ Moreover, it is assumed that the books are not exerting too much pressure among them, in their usual location on the shelf.



Figure 9. The “gripper” with the fingers specially designed.

Once the system is visually guided to achieve the location where approximately will be the book to grasp, the contact stage is initiated. Thus, the grasping algorithm for the contact stage requires three main steps:

1. Approaching to the book located by vision.
2. Looking for the left side of the book and inserting the left finger.
3. Looking for the right side of the book and inserting the right finger, while the left finger remains in its previous location.

As it was observed, all these steps are guided by means of force feedback. In fact, in Fig 10, it is summarized the signal processing from the force sensor output. Basically there are two main process: one that it is responsible to make the geometric transformations from sensory input associated to each axis, with the aim to adapt them to the robot system of reference; and other, enable to detect force patterns associated to some physical actions (e.g. a negative force in the Z direction means that the gripper is impacting with a book).

In summary, the signal processing, from force sensor in each one of the three axis, generates a force vector in the 3D space. By means of this vector, the algorithm computes the velocity vector associated to the spatial movement of the robot arm. The library, implemented in C, for the PA10 Arm control of movement, enable specify

the path associated to the gripper by means of a velocity vector

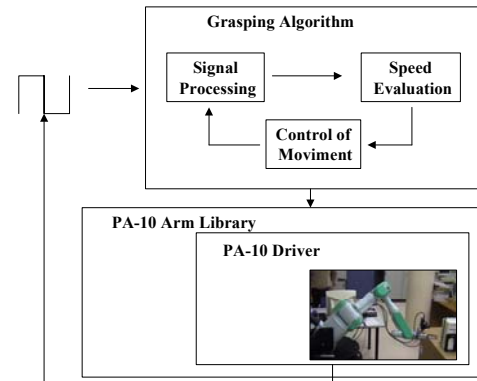


Figure 10. The grasping algorithm.

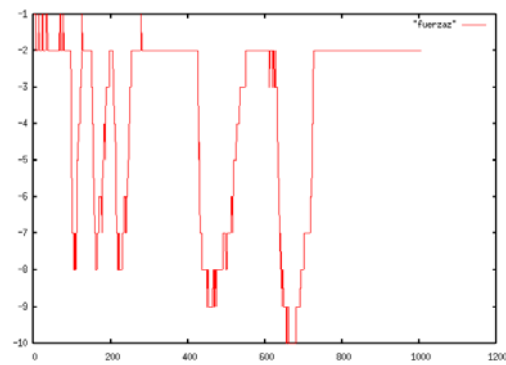


Figure 11. Evolution of forces applied along the Z axis during the grasping of a book.

With the aim to clarify the grasping process, the complete sequence of actions is shown in the following. As can be observed in Fig 11, there are three main regions connected with each one of the steps aforementioned for the grasping algorithm:

- (1) Approaching step related to the region between 0 and 80 approx., in abscissas. Here the gripper is moving towards the book, and finally impact with it, as is observed in the first pick for a value of 100 in abscissas.
- (2) Insertion of the left finger. The region between 220 and 400 approx., where no forces in Z are detected, indicate that the correct side of the book has been found, and in correspondance, the left finger can penetrate 2 cm without problems. In this location is fixed the left finger while the right is looking for the other side of the book.
- (3) Insertion of the right finger. Between 410 and 700, the forces indicate that the right finger is in contact with the surface of the book, but in 700 approx.,

the force disappears and between 700 and 1000 approx., the insertion of the right finger is accomplished.

5 Preliminary Results

The experimental setup, that can be appreciated in Fig 12, requires a mobile platform (Nomadic), a robot arm (Mitsubishi PA-10), and a camera in hand (MEGA-D Stereo Head) configuration. Moreover, an special design for the fingers in the parallel jaw griper, it was necessary, as observed previously in Fig 9. The system in action can be observed in Fig 12, where 4 main stages are highlighted:

1. The Image processing and Optical Character Recognition (OCR) [1], necessary to build the interface. This first step is in connection with the next one (2), because a scanner mechanism is needed to systematic process the different book labels until is found the one required by the user.
2. This step has the double mission to looking for the desired label by the user and, whether it was found making the approaching towards this book. All the time, in this non-contact step, the robot arm is vision guided.
3. Once the approaching stage is successful accomplished, become the grasping. Where the left finger of the parallel jaw griper is initially guided by vision to the left book border. After this, when the left finger is conveniently positioned, became the rest of the grasping process (i.e. contact stage), exclusively guided by means of force feedback.
4. Here the grasping process is successfully completed.

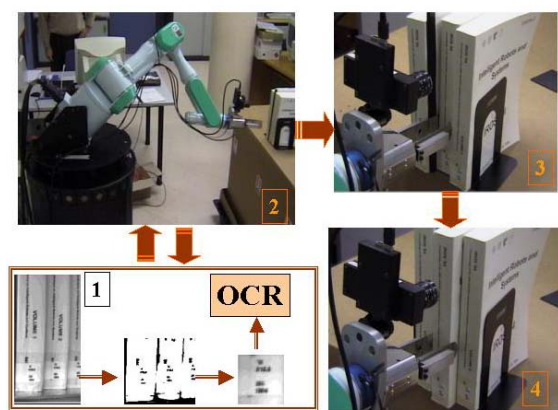


Figure 12. The system in action.

The first experiments developed with this system show a good performance both, in the quality results obtained and in the final consuming time for the complete

process. For instance, the time necessary to localize a label from an initial image is always less than 1 second. Moreover, the next image processing, including OCR, to identify the label, is always within the interval 5-10 seconds. On the other hand, related to the grasping process, a ratio of 5 seconds per 1 centimeter of book will be needed by means, before the book is grasped (e.g. a book of 4 cm wideness requires 20 sec). All these experiments have been developed by using a 266 MHz Pentium II Processor.

6 Concluding Remarks

These initial experiments are assuming some convenient situations (e.g. the arrangement of books in the shelf are in a vertical position, etc.). Thus, our short term objective will be to improve the performance and efficiency of our algorithms, meanwhile these assumptions are relaxed towards a more realistic scenario. Experiments showing the integration of a visual servoing module and grasping must be designed. Finally, as long term objectives, we are planning the complete integration of the modules necessary to scan among all the bookshelves, incorporating navigation capabilities, and testing the complete system in the real library scenario.

Acknowledgements

This paper describes research carried out at the Robotic Intelligence Laboratory of Universitat Jaume-I. Support for this research is provided in part by the MCyT under project DPI2001-3801, by the GV under projects GV01-244, and by the F. Caixa-Castelló under project P1-1B2001-28.

References

- [1] "Clara OCR", URL: www.claraocr.org
- [2] Sanz PJ, del Pobil AP, Iñesta JM, Recatalá G. "Vision-Guided Grasping of Unknown Objects for Service Robots". In Proc. IEEE Intl. Conf. on Robotics and Automation, 3018-3025, Leuven, Bélgica, 1998.
- [3] Suthakorn J, Lee S, Zhou Y, Thomas R, Choudhury S, Chirikjian GS. "A Robotic Library System for an Off-Site Shelving Facility". In Proc. IEEE Intl. Conf. on Robotics and Automation, 3589-3594, Washington, DC, USA, 2002.
- [4] Tetsuo Tomizawa, Akihisa Ohya and Shin'ichi Yuta. Object "Book Browsing System using an Autonomous Mobile Robot Teleoperated via the Internet". In Proc. of the IEEE Int. Conf. on Intelligent Robots and Systems (pp. 1284-1289). Lausanne. 2002.