iat

Sorin Mihai Grigorescu

Robust Machine Vision for Service Robotics

Shaker Verlag

Robust Machine Vision for Service Robotics

From the Faculty of Physics and Electrical Engineering
University of Bremen

for obtaining the academical degree of

Doktor-Ingenieur (Dr.-Ing)

approved dissertation

Dipl.-Ing. Sorin Mihai Grigorescu

Primary Supervisor:	Prof. Dr.-Ing. Axel Gräser
Secondary Supervisor:	Prof. Dr.-Ing. Walter Anheier
Submitted at:	01. October 2009
Date of Dissertation Colloquium:	20. January 2010

Publication Series of the Institute of Automation
University of Bremen

Series 7-Nr. 1

Sorin Mihai Grigorescu

Robust Machine Vision for Service Robotics

D 46 (Diss. Universität Bremen)

Shaker Verlag
Aachen 2010

To my mother
(Pentru mama mea)
Marmandiu Maria

Acknowledgements

I would like to thank Prof. Dr. Axel Gräser for his support and encouragement during my stay at The Institute of Automation (IAT) in Bremen, as well as for many helpful discussions. I would also like to thank my second reviewer, Prof. Dr. Walter Anheier and also to Prof. Dr. Udo Frese and Prof. Dr. Karl-Ludwig Krieger for showing interest to be on my dissertation committee.

A very special thank you goes to my friend Dr. Danijela Ristić-Durrant for her guidance and help during the years spent on my work. Without her the results presented in this thesis would probably never taken shape. Also, I wish to send my regards to Prof. Dr. Florin Moldoveanu who managed to give me during my years of study a part of his passion for engineering. I would like to show my gratitude to my friend Ing. Maria Johansson for her patience in correcting this thesis. My kind regards go to my colleagues, especially to Ing. Saravana Natarajan for helping me with evaluating the 3D object reconstruction results. I am also deeply thankful to my former students, Ing. Dennis Mronga and Ing. Björn Steinsträter, who invested a lot of their time in the development of the algorithms presented here. I thank also my IAT colleagues with whom I have worked with and wish them a good luck in their research activities and doctoral dissertations: Ing. Christos Fragkopoulos, Inf. Torsten Heyer, Ing. Uwe Lange, Ing. Adrian Leu, Ing. Leila Fotoohi, Ing. Tyaniu Wang and Ing. Thorsten Lüth. Many thanks to my former colleagues and friends, Dr. Oliver Prenzel, Dr. Dorin Aiteanu, Ing. Alin Brindusescu, Inf. Sorin Ivascu, Ing. Sai Krishna Vuppala and Inf. Chao Wang.

My deepest gratitude goes to my mother Maria Marmandiu and to my girlfriend Psi. Roxana Comnac. Thank you for supporting me all those years spent away from home and from you. Also, a big thank you goes to my Bremen family composed of Dr. Ioana Daraban and Chem. Daniel Daraban. I thank all my friends for the good moments that we had in the past years: Dr. Sandip Dhomse, Dr. Martyn Durrant, Psi. Daniela and Daniel Cristea, Psi. Costi Ciocoiu and Veronica Urjan.

Last but not least, I thank my mountain hiking friend that shaped a part of my character, Econ. Marian Huiban, who, during my research work, gave his life to the mountain.

Bremen, April 2010 Sorin M. Grigorescu

Abstract

In this thesis the vision architecture ROVIS (*RObust machine VIsion for Service robotics*) is suggested. The purpose of the architecture is to improve the robustness and accuracy of visual perceptual capabilities of service robotic systems. In comparison to traditional industrial robot vision where the working environment is predefined, service robots have to cope with variable illumination conditions and cluttered scenes. The key concept for robustness in this thesis is the inclusion of feedback structures within the image processing operations and between the components of ROVIS. Using this approach a consistent processing of visual data is achieved.

Specific for the suggested vision system are the novel methods used in two important areas of ROVIS: definition of an image ROI, on which further image processing algorithms are to be applied, and robust object recognition for reliable 3D object reconstruction. The ROI definition process, build around the well known "bottom-up top-down" framework, uses either pixel level information to construct a ROI bounding the object to be manipulated, or contextual knowledge from the working scene for bounding certain areas in the imaged environment. The object recognition and 3D reconstruction chain is developed for two cases: region and boundary based segmented objects. Since vision in ROVIS relies on image segmentation on each processing stage, that is image ROI definition and object recognition, robust segmentation methods had to be developed. As said before, the robustness of the proposed algorithms, and consequently of ROVIS, is represented by the inclusion of feedback mechanisms at image processing levels. The validation of the ROVIS system is performed through its integration in the overall control architecture of the service robot FRIEND. The performance of the proposed closed-loop vision methods is evaluated against their open-loop counterparts.

Kurzfassung

In der vorliegenden Dissertation wird das Bildverarbeitungsrahmenwerk ROVIS (*RObust machine VIsion for Service robotics*) vorgestellt. Dieses Rahmenwerk dient zur Verbesserung von Robustheit und Genauigkeit der visuell wahrnehmenden Fähigkeiten von Servicerobotiksystemen. Im Vergleich zu traditionellen Industrierobotern, bei denen die Arbeitsumgebung vordefiniert ist, müssen Serviceroboter variierende Beleuchtungsbedingungen und komplexe Umgebungen meistern. Das Schlüsselkonzept für die Robustheit in dieser Dissertation ist der Einsatz von Rückkopplungsstrukturen in den Bildverarbeitungsalgorithmen und zwischen den einzelnen ROVIS-Komponenten. Unter Verwendung dieses Ansatzes wird eine konsistente Verarbeitung der visuellen Daten erreicht.

Charakteristisch für das vorgeschlagene Bildverarbeitungssystem sind die neuartigen Methoden, die in zwei wichtigen Bereichen von ROVIS genutzt werden: die Definition von ROIs (Region Of Interest) im Bild, auf die dann weitere Bildverarbeitungsalgorithmen angewandt werden können, und die robuste Objekterkennung für zuverlässige 3D-Rekonstruktion. Das Verfahren zur Definition der ROI, das um das allgemein bekannte "bottom-up top-down" Rahmenwerk errichtet wurde, verwendet entweder Pixelinformationen zur Konstruktion einer ROI, die das interessierende Objekt enthält, oder kontextabhängige Erkenntnisse aus der Szene für die Begrenzung bestimmter Bereiche der visualisierten Umgebung. Die Objekterkennung und 3D-Rekonstruktion wurde für zwei Fälle entwickelt: bereichs- und kantenbasierte Erkennung von Objekten. Weil die Bildverarbeitung in ROVIS in jeder Verarbeitungsphase, d.h. bei der ROI-Definition und der Objekterkennung, auf Bildsegmentierung beruht, mussten robuste Segmentierungsalgorithmen entwickelt werden. Wie bereits erwähnt, wird die Robustheit der vorgestellten Verfahren und damit die Robustheit von ROVIS durch den Einsatz von Rückkopplungsstrukturen auf der Ebene der Bildverarbeitung erreicht. Eine Bestätigung der Güte von ROVIS ist durch die Integration im Steuerungsrahmenwerk des Serviceroboters FRIEND gegeben. Die Effizienz der vorgestellten visuellen Rückkopplungsmethoden wird durch einen Vergleich mit den zugehörigen Verfahren, die offene Regelkreise verwenden, bewertet.

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at http://dnb.d-nb.de.

Zugl.: Bremen, Univ., Diss., 2010

Publication Series of the INSTITUTE OF AUTOMATION, UNIVERSITY OF BREMEN:

1 Colloquium of Automation, Salzhausen
2 Automation
3 Robotics
4 Control Theory
5 Brain Computer Interface
6 Virtual and Augmented Reality
7 Image Processing

Printed in Germany.

ISBN 978-3-8322-9146-4
ISSN 1861-5457

Shaker Verlag GmbH • P.O. BOX 101818 • D-52018 Aachen
Phone: 0049/2407/9596-0 • Telefax: 0049/2407/9596-9
Internet: www.shaker.de • e-mail: info@shaker.de

Contents

1 Introduction

In humans, biological vision represents the transformation of visual sensation into visual perception [98]. The analogous computerized operation, also known as *computer vision*, deals with interpretation of digital images for the purpose of visual understanding of a scene by a machine, that is a computer. In comparison to computer vision, *machine vision* deals with the combination of different computer vision techniques and dedicated hardware for solving different tasks in fields like industrial manufacture, safety systems, control of Automated Guided Vehicles (AGVs), monitoring of agricultural production, medical image processing, artificial visual sensing for the blind or vision for robotic systems, also referred to as *robot vision*. The increased research in robot vision in the past years has spawned a large amount of systems and applications. From available literature, robot vision systems can be grouped according to the type of application they were designed for:

- *vision for manipulation* which represents the class of robot vision applications designed to detect objects that can be grasped by a dexterous manipulator;
- *vision in mobile robotics* characterized by the vision systems used for autonomous robot navigation and path following;
- *vision for mobile manipulation* representing a hybrid vision system designed for both robot navigation and dexterous manipulation.

Depending on the robot vision application, the visualized scene can be found either in an industrial environment, where position and orientation of objects is predefined and the illumination controlled, or, as the case of service robots, the imaged scene consists of typical human surroundings where objects are occluded and visualized in variable illumination conditions.

In this thesis, the problem of designing, improving and implementing service robotic vision systems is approached. Service robots represent a class of robots designed to operate semi- or fully autonomously to perform tasks useful to the well-being of humans and equipment, excluding manufacturing operations [114]. The applications of service robots range from simple domestic systems (e.g. vacuum [110] or automatic pool cleanears [112]) to entertainment [111] and social robots [115]. A special case of service robots which received large attention in last years are assistive systems designed to help disabled and elderly people. Such a robotic platform is FRIEND (*Functional Robot with dexterous arm and user-frIENdly interface for Disabled people*), a semi-autonomous service robot in its 3rd generation designed to support disabled people with impairments of their upper limbs in Activities of Daily Living (ADL) and professional life. The system consists of a seven

Degrees-of-Freedom (7-DoF) manipulator mounted on an electrical wheelchair. FRIEND is equipped with various sensors that provide intelligent perception of the environment needed for task execution support. One of these sensors is a stereo camera system which provides visual information regarding the system's environment. In particular, this thesis concerns the improvement of visual perceptual capabilities of the robotic system FRIEND for visual guided object grasping, a field of robotics in which a computer controls a manipulator's motion under visual guidance, much like people do in everyday life when reaching for objects. A key requirement in this field is the reliable recognition of objects in the robot's camera image, extraction of object features from the images and, based on the extracted features, subsequent correct object localization in a complex 3D (three dimensional) environment.

The main problem with service robotic systems such as FRIEND is that they have to operate in dynamic surroundings where the state of the environment is unpredictable and changes stochastically. Hence, two main problems have been encountered when developing image processing systems for service robotics: *unstructured environment* and *variable illumination conditions*. Such a scene can be easily noticed in everyday life: when a human is searching for an object he/she looks for it through a multitude of different other objects. Although this process is relatively simple for humans, its implementation on a machine has a high degree of complexity since a large amount of visual information is involved. A second major problem in robot vision is the wide spectrum of illumination conditions that appear during the on-line operation of the machine vision system. In a large number of vision applications one important attribute used in object recognition is the color property of objects. In case of the human visual system color is a result of the processing done by the brain and the retina which are able to determine the color of an object irrespective to the illuminant. The ability of the human visual system to compute color descriptors that stay constant even in variable illumination conditions is referred to as *color constancy* [26]. Although the color constancy ability is taken for granted in the human visual system this is not the case of machine vision.

Reliable object recognition and 3D reconstruction in robot vision is approached in this thesis through image segmentation, which represents the partitioning of a digital image into subregions suitable for further analysis. In literature, a number of object recognition methods that calculate an object's *Position and Orientation* (POSE) without the use of image segmentation have been proposed. One important category of such methods are based on the *Scale Invariant Feature Transform* (SIFT) [57]. Since then, a large number of SIFT based methods has spawned applications in various fields of computer vision, ranging from robotics [91] to medical image processing [27]. SIFT is a transformation for image features generation which are invariant to image translation, scaling, rotation and partially invariant to illumination changes and affine projection. The method can be used to generate a set of key features of an object which can further be used to relocate the object in an image. The key features, also called keypoints, are defined as maxima/minima of *Difference of Gaussians* (DoG) that occur at multiple scales from an image convolved with Gaussian filters at different scales [57]. Although the method is relatively robust with respect to occlusions and illumination conditions its major drawbacks are that firstly it needs an a priori model of the object to be recognized and secondly that the object has to be textured in order to calculate as many key features as possible. Also, the precision of the method is low if the object does not have a planar surface. These drawbacks motivate again the usage of image segmentation in robot vision where the precision of 3D object POSE estimation is crucial.

In Figure 1.1 the dependencies of object recognition and 3D reconstruction with respect to image segmentation are shown. The arrows in the figure represent the possible flow of information that may exist in a robot vision system where segmentation plays a central part. Image features, which provide object position in the 2D (two dimensional) image plane, extracted from binary segmented images, are used for recognizing object types and also to reconstruct their 3D shape in a virtual Cartesian space. Also, as it will be explained in Chapter 6.1.2, information regarding recognized objects can be used to improve segmentation quality, which directly influences precision of POSE estimation, that is the precision of detected 2D object feature points used for 3D reconstruction. Feature points are defined as key object points, obtained from segmented images, from which the object's 3D shape can be build. The type of object determined via 2D recognition influences the 3D reconstruction method in the sense that different feature points are extracted for different objects (e.g. for a bottle, its top neck and bottom are used as feature points, as for a book, feature points are represented by the four corners of the book).

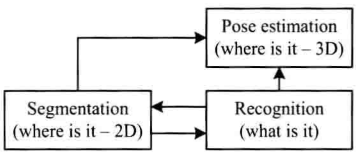

Figure 1.1: Object recognition and 3D reconstruction in robot vision.

In this thesis, the novel robot vision system ROVIS (*RObust machine VIsion for Service robotics*) is suggested as a contribution to reliable and robust object recognition and 3D reconstruction for the purpose of manipulation motion planning and object grasping in unstructured environments with variable illumination conditions. In ROVIS, robustness must be understood as the capability of the system to adapt to varying operational conditions and is realized through the inclusion of feedback mechanisms at image processing level and also between different hardware and software components of the vision framework. A core part of the proposed system is the closed-loop calculation of an image Region of Interest (ROI) on which vision methods are to be applied. By using a ROI, the performance of object recognition and reconstruction can be improved since the scene complexity is reduced.

1.1 Related work and contribution of the thesis

According to the classification of robot vision applications made in introduction, a number of representative state-of-the art image processing architectures, related to the ROVIS system proposed in this thesis, are summarized below.

Vision for manipulation

Localization of objects for manipulative purposes has been treated in a relative large number of publications. A framework for object manipulation in domestic environments, based on visual robot control and a cognitive vision architecture, is proposed in [46, 47, 48]. The experimental platform for the framework is a Nomadic Technologies XR4000, equipped with a Puma 560 arm for object manipulation and a stereo camera set for scene recognition. The approach used in detecting objects of interest for grasping and manipulation is divided into localization of known and unknown objects in cluttered environments, using the SIFT method. The complex problem of unstructured environments is also discussed in [20] with respect to visual guided object grasping. In comparison to the mentioned work, in this thesis the problem of object recognition is treated from the point of view of improving the classical open-loop image processing chain and not from applying complex vision algorithms for object detection.

Investigations regarding the architectural aspects of robot vision for manipulative tasks can be found in [24], where a control framework for "hand-eye" manipulation, which manages complexity of tasks through the composition of a few simple primitives, is proposed. In [97], an extensive work regarding visual perception and control of robotic manipulation is given. Here, data processed by the vision system, based on 3D model shapes, is enhanced with shape recovery information acquired from robust light stripe scanning. In ROVIS, instead of using 3D primitive shape models for object reconstruction, 2D shape descriptions of objects are used for recognition and 2D feature points calculation.

In recent years, the application of vision based object manipulation has been extensively applied in the field of assistive and rehabilitation robots, as the case of the vision system proposed in this thesis. AVISO (*Assistance by VIsion for Seizure of Objects*), detailed in [52, 25], is composed of a MANUS arm and a stereo "eye-in-hand" camera system mounted on the end-effector of the arm. The AVISO vision system relies on the end-user (patient) to cope with the separation of the object to be grasped from the background and also from the other objects. This process is achieved through the user by manually moving the MANUS manipulator at a distance of approximatively ten centimeters to the object. The arm movement is controlled using a human-machine interface which displays in real-time images captured from the stereoscopic "eye-in-hand" camera. Once the object is centered on the displayed image, the user has to manually draw an image ROI bounding the object to be grasped. Finally, the object is automatically separated from the background by calculating interest points with the Harris and Stephens [33] feature detector. Using the epipolar geometry constraints, a 3D point representing the object grasping point is calculated. Further, the manipulator arm approaches and grasps the object in a visual control manner. Although the AVISO method is reliable in unstructured environments, it implies a relatively large amount of user interaction, which can be tiring for the patient. An alternative method for AVISO, which uses less user interaction, is proposed in [25], where the user must only select a point from the object to be grasped on an image captured from the "eye-in-hand" camera. Using a second fixed camera present in the scene and epipolar geometry the system can approach and grasp the object. This method, although requiring less user interaction, it relies on the second camera present in the scene. In this thesis, user interaction is used for defining an interest point only once, hence limiting tiring interaction tasks. The defined interest point acts as a starting point for automatic adjustment of the image ROI, on which the object is automatically localized.

Similar to AVISO [52, 25], the AMOS (*Assistive MObile robot System*) vision system [96] also uses a stereoscopic "eye-in-hand" camera mounted on a manipulator arm and a *Shared Responsibility* software architecture which involves the user in the working scenarios of the robot. If the object to be grasped is out of the range of the manipulator, the position of AMOS is changed in order to get it closer to the object. In both systems presented above, AVISO and AMOS, "eye-in-hand" cameras are used. In comparison to that, in ROVIS visual information is obtained from a global, fixed, stereo camera system. The advantage of using a global camera over an "eye-in-hand" one is that it provides a global description of the imaged scene, which is more appropriate for manipulator motion planing with obstacles avoidance.

In [56] the visual control of the MANUS arm using the SIFT [59] algorithm is proposed. As said before, the disadvantage of this approach is the need of a 3D model of the object to be grasped (the SIFT-keypoints) and the fact that the SIFT algorithm provides reliable results only for planar textured objects. Recently, SIFT was used in a shared-controlled architecture to simulate ADL tasks, also considered in this thesis, using a visual controlled assistive robot arm [44].

Vision in mobile robotics

In mobile robotics, vision is commonly used for autonomous navigation control, indoor or outdoor, of a mobile robotic platform. Although the vision system proposed in this thesis has as target object recognition for manipulation, concepts from vision for mobile robotics are presented here as applied to general robot vision.

A survey of color learning algorithms and illumination invariance in mobile robotics is given in [94]. The algorithms are presented from the perspective of autonomous mobile robot navigation with stress on inter-dependencies between components and high-level action planning. As an alternative to the color segmentation methods presented in the survey, in this thesis, robustness of color segmentation against varying illumination is achieved through feedback adaptation of the image processing parameters.

The vision system proposed in [13] is one of the first ones to approach the robot vision problem from an image processing architectural point of view. The SRIPPs (*Structured Reactive Image Processing Plans*) image processing system is designed for the robot control architecture FAXBOT of the RHINO mobile robotic platform. The sequence of vision algorithms is built around an image processing pipeline. As a comparison, the architecture of ROVIS is modeled using the *Unified Modeling Language* (UML) [65]. Through the use of UML a better transparency of the system is achieved, as also a simplified development process since the structure of the computer programs that make up ROVIS is implemented graphically.

An interesting "visual attention" computer vision architecture developed for mobile robots is VOCUS (*Visual Object detection with a CompUtational attention System*) [28, 29]. The designed system is based on the cognitive capabilities and neuro-physiology of the human brain. For the recognition of objects two approaches are used:

- "bottom-up" attention, when no a priori information regarding the visualized scene exists;
- "top-down" approach, when a priori information about the objects to be recognized in the scene exists.

The "bottom-up top-down" framework is also used in this thesis for building two methods for image ROI definition. Attention vision architectures have also been studied in [75] with the purpose of optimizing the sensorimotor behavior of a mobile robot.

Vision for mobile manipulation

The systems designed for both navigation and object manipulation are mobile robotic platforms equipped with redundant manipulators. Such hybrid systems, like the mobile robot JL-2 [107] used for field operations, rely on vision to calculate the robot's moving path and also recognize objects to be manipulated.

In recent years, robotic perception in domestic environments has been treated in a large number of publications [12, 47]. The UJI librarian robot [22] 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 the vision system in [22], ROVIS aims at the recognition and 3D reconstruction of all types of books, placed in cluttered environments.

Care-O-Bot® represents a series of mobile robots designed to assist people in daily life activities [32, 89]. They were developed at Fraunhofer IPA [1] since 1998, currently reaching its 3rd generation. The object recognition system of Care-O-Bot uses a camera sensor and a 3D laser scanner for reconstructing the 3D representation of objects of interest. The objects are taught beforehand to the robot using model images. Also, a laser scanner is used for planning a collision free trajectory of the 7-DoF manipulator arm used in grasping

[1] Fraunhofer-Institut für Produktionstechnik und Automatisierung

and manipulating objects. In ROVIS, the goal is to develop a robust vision system based only on image data from a global stereo camera with the purpose of extracting as much visual information as possible.

In [87, 88], it has been investigated how a mobile robot can acquire an environment object model of a kitchen and more generally of human living environments. For this purpose, range sensing information acquired from a laser scanner, in form of 3D point clouds, has been used. 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.

A different approach in researching mobile manipulation is found in the BIRON robot (*Bielefeld Robot Companion*), where the focus is on manipulative actions for human-robot interaction [55]. The topic of human-robot interaction, also treated in [14, 15, 52, 96], plays a very important role in robotics, generally concerning safety and particularly regarding recognition and interpretation of human gestures. In this thesis, human-machine interaction is treated as a tool used for sending input commands to the robotic system.

The contributions of this thesis, with respect to the state-of-the art systems presented above, are both theoretical and practical, as summarized below.

The thesis considers the design and implementation of a robot vision system with improved visual perceptual capabilities for the purpose of manipulator path planning and object grasping [73, 72]. The focus of the proposed vision system ROVIS is specifically related to service robotic platforms, namely it treats the various form of interconnections between system components and also the connection to the overall control architecture of the robot. These conceptual elements represent a crucial factor in the well operation of a robotic system [46, 77].

The thesis represents a contribution in the field of feedback control in image processing [83, 70, 66, 49]. Because of the complex environment where ROVIS operates, its robustness with respect to external influences is critical. The role of including feedback structures at image processing level is to improve this robustness. The region based color segmentation algorithm proposed in this thesis represents further research in the field of region segmentation, as a sequel to the proved closed-loop gray level region segmentation from [83]. The case of optimal boundary segmentation has been approached with investigating a new feedback quality measure derived from feature extraction level. The objective of both segmentation methods is to reliably extract 2D object feature points needed for 3D reconstruction. As said before, the investigation of feedback mechanisms for ROVIS are intended for the improvement of the overall robustness of the vision system following the principle of decentralized control [40], where the robustness of a complex system is not achieved by a complex control mechanism, but by controlling subcomponents of it, thus ensuring overall system robustness.

One other important aspect in designing ROVIS is its performance evaluation with respect to traditional vision architectures, where image processing is performed in an