Integrated System for Stereoscopic Cognitive Vision, Localization, Mapping, and Communication with a Mobile Service Robot

Cătălin Buiu

POLITEHNICA University of Bucharest Department of Automatic Control and Systems Engineering Spl. Independentei 313, 060042 Bucharest, Romania E-mail: cbuiu@ics.pub.ro

> **Abstract:** This paper describes a stereo-vision-based mobile robot that can navigate and explore its environment autonomously and safely and simultaneously building a tridimensional virtual map of the environment. The control strategy is rule-based and the interaction with robot is done via Bluetooth. The stereoscopic vision allows the robot to recognize objects and to determine the distance to the analyzed objects. The robot is able to generate and simultaneously update a full colour 3D map of the environment that is being explored. The position and type of each detected and recognized object is marked in this 3D map. Furthermore, the robot will be able to use a gripper in order to collect detected objects and carry them to dedicated collecting bins, and so will be able to work in commercial waste cleanup applications. This application represents a successful integration of computers, control and communication techniques in mobile service robot control.

> **Keywords:** control, communication, localization, mapping, mobile robot, stereoscopic vision, virtual reality

1 Introduction

More than 7 million robots will be sold from 2005 to 2008 according to estimations of the International Robotics Federation and of the Economic Commission for Europe of the United Nations. Until 2010 a robust increase of 4% per year in the number of robots is estimated. Many of these are service robots which are used to assist or even to replace humans in tedious, dull, dangerous or repetitive tasks. The same sources estimate that by 2010, service robots will be able to fully assist elder people and people with disabilities, will extinct fires, will explore industrial pipes and more [1].

In ecological applications, service robots are used to collect waste and dangerous items in indoor and outdoor environments. For doing that, the robots must be able not only to perceive and act upon the environment by using a wide range of sensors and actuators, but also to manifest human-like cognitive abilities, such as to localize themselves, to recognize and classify objects, to generate maps of the environment, to learn from experience, to interact in a natural way with humans and other robots or to develop physical and cognitive abilities in a kind of developmental process similar to humans. It is often the case that a team of robots is asked to fulfill such a task. There are already a lot of interesting results obtained in collective robotics, see for example [2] where coordinated control based on artificial vision is investigated and [3] where decentralized formation control of mobile robots with limited sensing is addressed.

The problem of Simultaneous Localization and Mapping (SLAM) consists of estimating concurrently the robot's position and generating a map of its surrounding environment. This is an essential skill for a mobile robot but to this day it has eluded complete and robust solutions because noisy robot dynamics and sensors make solving SLAM a difficult task.

SLAM has been widely used for navigation and typically makes use of laser range-finders or sonars. An advantage of using stereo vision over laser range-finders is the ability to detect obstacles at different heights. The solution presented in [4] is based on learning maps of 3D point-landmarks whose location is estimated using correlation-based stereo and identification is performed using their appearance in images using the Scale Invariant Feature Transform (SIFT) [5]. The authors derive an estimate of the robot's motion from sparse visual measurements using stereo vision and multiple view geometry techniques [6] known in robotics as visual odometry [7], [8], [9].

A number of professional stereo vision systems and related software systems have been developed and have found a number of interesting applications in various domains, from the control of industrial manipulators for assembly and pick-and-place operations, material handling, collision warning and obstacle detection in robotics, people-tracking, environment modeling, autonomous guidance of corn harvesters, digitizing books, to medical applications, such as ophthalmic diagnostics, IR mammography, and robotic laparoscopy.

Stereo vision for navigation has a long history and is frequently exploited for autonomous navigation, but has limitations in terms of its density and accuracy in the far field [10]. If landmarks can be placed in the field of view of the camera, the location of a vehicle can be determined by means of stereo vision [11], and if a solid model of the target object is available, a robotic manipulator will have at its disposal a modeled environment for automatic tasks [12]. In [13] it is presented a stereoscopic vision system for a Khepera miniature robot. The vision system performs objects detection by using the stereo disparity and stereo correspondence.

An adaptive panoramic stereo vision approach for localizing 3D moving objects has been developed in the Department of Computer Science at the University of Massachusetts at Amherst. In the adaptive stereo model, the sensor geometry can be controlled to manage the precision of the resulting virtual stereo system [14]. Other indoor and outdoor stereo vision systems have been developed and tested with satisfactory results and some drawbacks, see [15], [16] and [17].

A novel optical system allows the capture of a pair of short, wide stereo images from a single camera, which are then processed to detect vertical edges and infer obstacle positions and locations within the planar field of view, providing real-time obstacle detection [18].

Very few applications concern the problem of waste collecting service robots acting indoor. The application reported in this paper is part of the bigger ReMaster research project currently under development at the Autonomous Robotics Lab of the POLITEHNICA University of Bucharest, Romania. This project concerns the development of a commercial cognitive service robot to be used in waste cleanup in office buildings. A first prototype (ReMaster One) has been built. Related details on the structure of the prototype and of its cognitive vision system using a monocular vision system are given in [19] and [20]. The acquired expertise has been used to propose the structure and to design a stereoscopic cognitive vision system which is detailed in [21].

The aim of this paper is to present the current phase in the ReMaster project which consists in the design and implementation of an integrated system for stereoscopic cognitive vision, localization, mapping, and communication with the robot. The goals of this system is to allow the robot to recognize and classify various objects and to determine the distance to the objects. Combined with the self-localization ability of the robot, this allows the absolute position of detected objects to be determined and marked on a tridimensional map of the working space that is continously updated. So, stereo vision and SLAM are integrating in order to create a map of the environment, without using any landmarks. The realization of this vision based mapping is the main contribution of this paper.

The paper is structured as follows. Section 2 gives an overview of the system architecture, and of the robot control and communication system, while Section 3 presents the realization of the stereoscopic cognitive vision system. Section 4 describes the way in which the robot is able to generate, maintain and update a tridimensional virtual map of the environment that is being explored. The last section of the paper presents conclusions and some directions for further research and developments.

2 Robot Control and Communication System

The integrated system was implemented on a Koala robot (Fig. 1) which is a mid-size robot designed for real-world applications and capable of carrying larger accessories. It has been chosen for this application, as Koala has the functionality necessary for use in practical applications (like sophisticated battery management), and rides on 6 wheels for indoor and all-terrain operation. It has 16 distance sensors and can be controlled via Bluetooth.



Figure 1: Koala mobile robot (www.k-team.com)

Two commercial webcameras have been mounted on top Koala at the same level (Fig. 2), at a distance of 95 mm between them, and at a height of 170 mm. The cameras are inclined at 10 degrees, and have CMOS 1.3 megapixel sensors (1280*960 pixels images and 640*480 videos), manual focus, and a focal distance of 1/4.8 mm.

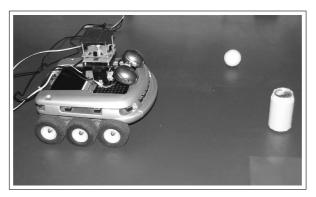


Figure 2: Stereoscopic vision system on-board Koala

On the robot side, there is a Bluetooth 333s module installed and directly connected to the serial interface of the robot. The control program runs on a separate laptop which communicates with the robot by using a Bluetooth connection. The robot can localize itself by using a dedicated and redundant system consisting of a beacon and two transponders ([19], [21]) and an odometry algorithm. The robot is able to navigate in indoor environments consisting of walls and various objects, such as empty cans and bottles.

The control program is implemented in Matlab and is based on simple control rules which allow an obstacle avoidance and waste finding behavior (Fig. 3). The robot will move forward and will be able of a safe navigation and detection of objects in the workspace (Fig. 4). After detecting an obstacle, the robot will stop and action according to the type of the obstacle. If wall, the obstacle will be avoided and the robot will resume moving. If not wall, using the distance sensors on-board the robot, the system will compute the distance to the object and the corresponding angle. Using these two measurements, the system will compute the absolute position of the detected object and will compare this with the stored coordinates of previously detected objects. If the object is new (its absolute position is not in the database), it will get more attention from the robot which will turn so that it is facing the object (Fig. 5).

Now, the system is ready to acquire stereoscopic images of the object. The images are processed and analyzed as explained in the next section. The distance to the object is determined and based on

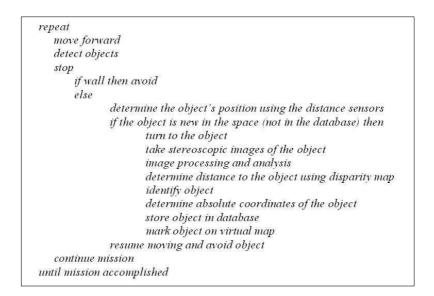


Figure 3: Control algorithm

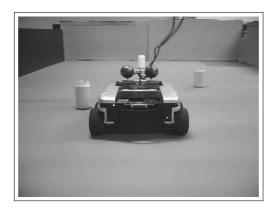


Figure 4: Robot navigating in a test environment



Figure 5: Robot turning to a detected object

the absolute position of the robot, the absolute position of the object is also determined. Then the robot resumes its movement in the workspace after storing the type and coordinates of the object in the database and after having marked its position in the tridimensional map which will be detailed later.

3 Stereoscopic Cognitive Vision system

Stereoscopy is a technique for infering the 3D position of objects from two or more simultaneously views of the scene. Reconstruction of the world seen through stereo cameras can be divided in two steps. First, the correspondence problem means that for every point in one image to find out the correspondent point on the other and compute the disparity of these points. This disparity correlates to distance, and the higher disparity of object pixel means that the object is closer to the cameras. Secondly, there is the triangulation step. Given the disparity map, the focal distance of the two cameras and the geometry of the stereo setting (relative position and orientation of the cameras) compute the (X,Y,Z) coordinates of all points in the images. The system presented in this paper solves both steps as will be described below.

Key advantages of camera based systems include: they offer minimally complex solutions, have very low costs, they are entirely solid state, and colour information can be easily acquired at the same time as range data, helping to build realistic full colour 3D models of the environment. All these advantages are exploited in the application reported in this paper.

Stereo vision provides realtime, full-field distance information, and is useful in many applications in a wide variety of fields, including robotics. There is a number of dedicated software packages, such as Small Vision System for realtime stereo analysis from SRI's Artificial Intelligence Center. Sentience is a volumetric perception system for mobile robots and uses webcam-based stereoscopic vision to generate depth maps, and from these create colour 3D voxel models of the environment for obstacle avoidance, navigation and object recognition purposes.

A "cognitive vision system" is defined in [22] as a system that uses visual information to achieve: recognition and categorization of objects, structures and events, learning and adaptation, memory and representation of knowledge, control and attention. For example, a cognitive monocular vision system for a mobile robot using a CMUcam2+ camera is presented in [20].

The visual system's architecture is presented in Fig. 6. All the visual information processing is done on the same separate laptop. Given the disparity map, the focal distance of the two cameras and the geometry of the stereo setting (relative position and orientation of the cameras), the system is able to compute the coordinates of all points in the images. The distance to the object is used to determine the absolute position of the object which is marked on the 3D map.

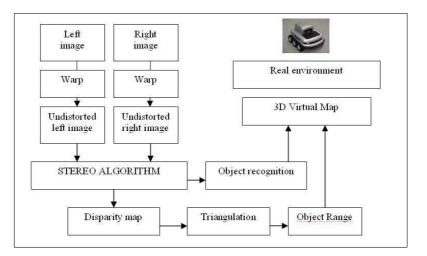


Figure 6: Stereovision system's architecture

Screenshots from our application that present two stereoscopic images of a detected object are given in Fig. 7. These images will be further processed.

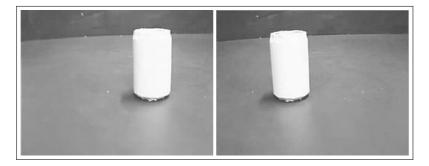


Figure 7: Stereoscopic images of a detected object (left and right hand camera)

Now the images contain relevant data that will be brought in such a form that contours can be extracted. The images are binarized by extracting colour channels corresponding to the colour of detected objects (yellow, in our case). Then, dilatation and errosion filters are applied to the images (see Fig. 8). Then the images are segmented and objects detected (Fig. 9).

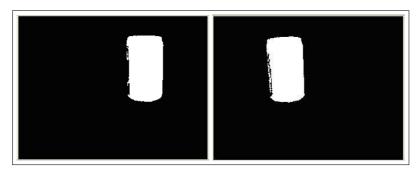


Figure 8: Extraction of yellow colour channel and application of dilatation and errosion filters (left and right hand image)

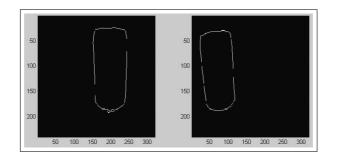


Figure 9: Detected contour for the object in the image

Using simple scalar descriptors, such as area and perimeter, the detected object is recognized and classified as a can, in our case.

4 Generation and Update of a Tridimensional Map of the Environment

Simultaneous Localization and Mapping (SLAM) is an essential capability for mobile robots exploring unknown environments. The robot presented in this paper is using a dedicated self-localization

system based on the use of a Beacon unit and two Transponders (master and slave) [21]. The two transponder units are fixed, while the beacon unit is installed on the robot. Half-duplex bidirectional communication between beacon and transponders is realized by using infrared light and ultrasounds. The system is using a ATmega8 microcontroller with 16MIPS at 16MHz. The localization of the robot is realized by triangulation of the distances to the two transponders. More, odometry algorithms contribute to a more precise localization of the robot in the working space.

The system is able to generate a a virtual map of the explored environment, in which the space, the robot and the objects are modeled as VRML (Virtual Reality Modeling Language) objects. VRML is a standard file format for representing tridimensional interactive vector graphics. It also enables the integration of interactive 3D graphics into the Web.

By using the Virtual Reality Toolbox from Matlab, the system will generate realistic 3D views of the working space and the robot (Fig. 10), and objects (Fig. 11, in which a question mark means an unknown object).



Figure 10: VRML models of the working space and robot

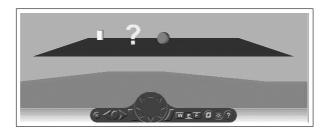


Figure 11: VRML models of objects

The robot will explore the working space according to the control strategy presented above and simultaneously will update the tridimensional map. After an object is detected and the robot turns towards it, both cameras are taking images of the scene. Further, the images are analyzed and the object recognized. The absolute position of the object is also determined and the object marked on the map (Fig. 12). Then the robot resumes its movement in the workspace and associated activities: navigation, search, classification and localization of objects.

The test results show a good and robust functioning of the stereovision system, and although the processing times are not low, this can be improved by the use of an embedded PC with more computing power.

5 Summary and Conclusions

The main research thrust of this paper has been to demonstrate that an integrated system for communication, control, localization and mapping using stereoscopic vision and 3D maps can be designed and implemented for a mobile service robot which will collect waste in indoor environments. This integrated system will be transferred to a more powerful version of the first prototype (ReMaster One) of the commercial waste cleanup robot that is the final aim of the ReMaster project. The new robot will have



Figure 12: Robot taking pictures of an object, recognizing it as a can, and marking it on the map

a gripper such that the detected objects can be grasped and carried to dedicated bins. Future efforts will address the design and implementation of new navigation strategies based on fuzzy logic. Image processing algorithms based on cellular neural networks are currently under investigation and implementation. More research will be done in what regards the interactive aspects of the robotic system so that the robot will be able to interact with humans and other robots in a natural way.

Acknowledgements

We acknowledge the support of the Romanian Government through the Excellence Research program (Contract 83-CEEX-II-03/31.07.2006) and the work of Cristian Ionita and Laura Antochi to the development of the stereoscopic vision system and virtual map.

Bibliography

- [1] C. Buiu (Editor), Cognitive Robots (in Romanian), Editura Universitara, 2008.
- [2] C. M. Soria, R. Carelli, R., J. M. Ibarra Zannatha, Coordinated Control Of Mobile Robots Based On Artificial Vision, International Journal of Computers, Communications, and Control, Vol. I (2006), No. 2, pp. 85-94.
- [3] K.D. Do, Bounded Controllers for Decentralized Formation Control of Mobile Robots with Limited Sensing, International Journal of Computers, Communications, and Control, Vol. II (2007), No. 4, pp. 340-354.
- [4] P. Elinas, R. Sim, J. J. Little, SigmaSLAM: Stereo vision SLAM using the Rao-Blackwellised particle filter and a novel mixture proposal distribution, In Proc. of the IEEE Int. Conf. on Robotics and Automation (ICRA), Florida, USA, 2006.
- [5] D. G. Lowe, Object recognition from local scale-invariant features, In Int. Conf. on Computer Vision, Corfu, Greece, September 1999, pp. 1150-1157.
- [6] R. Hartley, A. Zisserman, Multiple View Geometry in Computer Vision, Cambridge, UK: Cambridge Univ. Pr., 2000.
- [7] D. Nister, O. Naroditsky, J. Bergen, Visual odometry, In Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2004), 2004, pp. 652-659.

- [8] M. Agrawal, K. Konolige, Rough terrain visual odometry, In Proceedings of the International Conference on Advanced Robotics (ICAR), August 2007.
- [9] K. Konolige, M. Agrawal, Frame-frame matching for realtime consistent visual mapping, In Proceedings of IEEE International Conference on Robotics and Automation (ICRA), April 2007.
- [10] M. J. Procopio, T. Strohmann, A. R. Bates, G. Grudic, Jane Mulligan, Using Binary Classifiers to Augment Stereo Vision for Enhanced Autonomous Robot Navigation, University of Colorado at Boulder Technical Report CU-CS-1027-07, April 2007
- [11] Wang, L. K., S. Hsieh, E. C. Hsueh, F. Hsaio, K. Huang, Complete pose determination for low altitude unmanned aerial vehicle using stereo vision, In Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2005), pp. 108 - 113.
- [12] Lee, S., D. Jang, E. Kim, S. Hong, J. Han, A real-time 3D workspace modeling with stereo camera, In Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2005), pp. 2140 - 2147.
- [13] T. Chinapirom, U. Witkowski, R. Ulrich, Stereoscopic Camera for Autonomous Mini-Robots Applied in KheperaSot League, Research Report, University of Paderborn, Germania, 2007
- [14] D. R. Karuppiah, Z. Zhu, P. Shenoy, E. M. Riseman, A fault-tolerant distributed vision system architecture for object tracking in a smart room, In B. Schiele and G. Sagerer (Eds.), Springer Lecture Notes in Computer Science 2095, pp 201-219, 2007
- [15] S. Florczyk, Robot Vision: Video-based Indoor Exploration with Autonomous and Mobile Robots, Weinheim: Wiley-VCH, 2005.
- [16] M. F. Ahmed, Development of a Stereo Vision system for Outdoor Mobile Robots, M.S. thesis, University of Florida, 2006.
- [17] F. Rovira-Más, S. Han, J. Wei, J. F. Reid, Autonomous Guidance of a Corn Harvester using Stereo Vision, Agricultural Engineering International: the CIGR Ejournal, Manuscript ATOE 07 013, Vol. IX. July, 2007.
- [18] W. Lovegrove, B. Brame, Single-camera stereo vision for obstacle detection in mobile robots, In Intelligent Robots and Computer Vision XXV: Algorithms, Techniques, and Active Vision., Proceedings of the SPIE, Volume 6764, pp. 67640T, 2007
- [19] C. Buiu, F. Cazan, R. Ciurlea, Developing of a Service Robot to Recognize and Sort Waste, In: 16th International Conference on Control Systems and Computer Science, pp. 298-303. POLITEHNICA Press, Bucharest, 2007.
- [20] Ana Pavel, C. Vasile, C. Buiu, Cognitive Vision System for an Ecological Mobile Robot, In Proceedings of SINTES 13, The International Symposium on System Theory, Automation, Robotics, Computers, Informatics, Electronics and Instrumentation, pp. 267-272, Universitaria Press, Craiova, 2007.
- [21] C. Buiu, Design and development of a waste cleanup service robot, In Proceedings of the First International EUROBOT Conference, Heidelberg, pp. 194-202, 2008.
- [22] A.G. Cohn, D. Magee, A. Galata, D. Hogg, S. Hazarika, Towards an architecture for cognitive vision using qualitative spatio-temporal representations and abduction, In C. Freksa, W. Brauer, C. Habel, K.F. Wender (editors), Spatial Cognition III, Routes and Navigation, Human Memory and Learning, Spatial Representation and Spatial Learning, pp. 232-248, Springer-Verlag, 2003.