Vision-based systems for autonomous driving and mobile robots navigation

Ubiquitous computing seminar FS2014
Student report

Lukas Häfliger
11-916-376
haelukas@student.ethz.ch

ABSTRACT

The following paper, compares different vision based systems used in mobile robots navigation and in autonomous driving. The first part focuses on the differences between different areas of application such as space, underwater and air. In the second part we mainly focus on the autonomous driving part where different approaches, for the different problems that arise, are presented.

INTRODUCTION

There are many areas where autonomously operating robots can be an asset. One can consider for example underwater deep sea cable inspection. Such a cable may have a depth of up to 8000 meters so sending down a human in a submarine is not considered practical. On the other hand using a robot which is controlled by humans is very expensive since it must have a cable connection, so a boat must be close at all times. Since this is a task which arises quite regularly the main focus lies on developing an autonomously operating robot. Now especially robots which operate in space have certain constraints such as extreme reliability and autonomous obstacle detection and path planning. These systems suffer in a similar way from difficult communication with such a device due to the great distances. When talking about Curiosity[1] one must take into account that the transmission delay is approximately 7 minutes. So a human controlling such a robot is not practical.

NAVIGATION ON GROUND

Systems which operate on earth on the ground have the least constraints of any other systems. These systems are neither bound by their size, their power consumption nor their weight. So these systems are pretty reliable now and work fairly well as seen in [4]. To discuss the outdoor systems, the STANLEY robot is presented. In the second part approaches for indoor systems are presented. The STANLEY robot was able to drive fully autonomously through a 175 mile desert course. STANLEY is equipped with five SICK laser range finders which measure the terrain ahead up to 25 meters. Additionally a color camera is mounted on top of STANLEY to extend the measurements from the laser range finders. Finally there are also two RADAR[2] sensors. Now one may ask where the current research is focusing on. The used systems in STANLEY are pretty big and use about 500W so they are relatively expensive. If one could build a system which only uses multiple cameras, such a system would be extremely cheap.

Outdoor approaches

In systems which operate on ground, common system are composed of one or more laser range finders and multiple cameras. These systems typically build occupancy grid using machine learning algorithms. For path planning GPS[3] can be used but GPS outages must be taken into account.

Laser range finders A laser range finder works the same way a RADAR system does. It sends out a laser beam and measures the time it takes the laser beam to be reflected and sent back. Out of this data and the direction the beam was sent, such a system can reconstruct an exact 3D model of its surrounding.

Range extension with cameras In STANLEY a color camera was used to extend the range of the laser range finders. The motivation behind this approach was, that the laser range finders could only perceive the terrain up to 25m ahead which would prevent STANLEY from driving more than 25mph. To extend the range, the classification of the laser range finders is used as a learning set for the machine learning algorithm. With this algorithm the range is the improved based on the color cameras image. To emphasis the importance of this system one may consider figure 1. In these images, the machine learning algorithm is shown in action. In the first two images, the pavement road is classified as drivable by the laser range finders. As STANLEY gets the command to drive on the grass, more and more of it is classified as drivable, so one sees that the machine learning algorithm quickly adapts to this change.

Occupancy grids The terrain ahead of the robot is divided into cells. Each of these cells are then either classified as drivable or not. In STANLEY they used a classification based

[1] Curiosity is a car-sized robotic rover exploring Gale Crater on Mars as part of NASA's Mars Science Laboratory mission (MSL)

[2] Radio Detection and Ranging

[3] Global positioning system

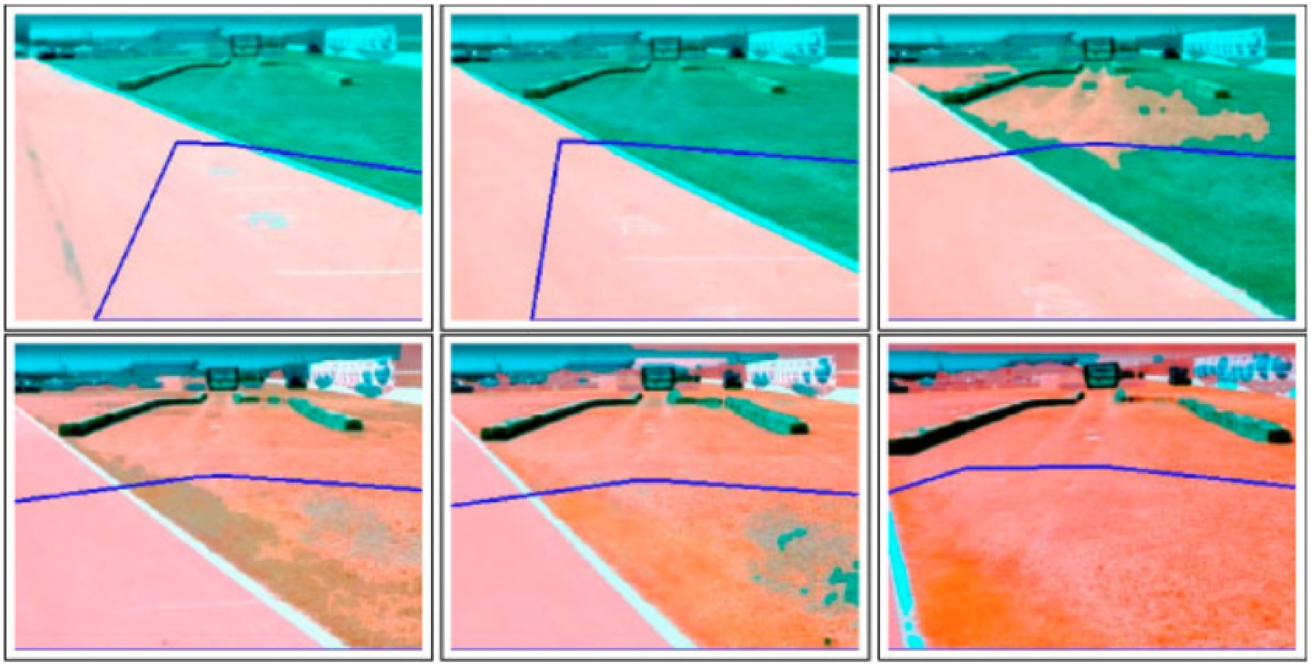


Figure 1: These images illustrate the rapid adaptation of Stanley's computer vision routines. When the laser predominately screend the paved surface, the grass is not classified as drivable. As Stanley moves into the grass area, the classification changes. This sequence of images also illustrate why the vision result should not be used for steering decisions, in that the grass area is clearly drivable, yet Stanley is unable to detect this from distance.[4][10]

on the vertical distance of adjacent cells. Then a simple threshold is used for classification.

Indoor approaches

To provide some basic overview a few terms need to be specified. When talking about indoor navigation map-based, map-building and mapless systems can be used.

- Map-based: These systems are provided with a map but due to a lack of GPS they need to be able to localize themselves within this map. Additionally we have moving obstacles like humans which cannot be evaded by just using the map, so these systems still need some sort of obstacle detection
- Map-building: These systems build their own map in a first step and use it later to navigate through the environment.
- Mapless: These systems just react to their surroundings.

If one summarized the most important approaches for all these systems, the following techniques would be mentioned:

- Force fields
- Occupancy grid
- Stereo 3D reconstruction
- Optical Flow
- Appearance based

Now all these approaches are shortly explained.

Force fields Force fields are used in the map-based approach. In this approach, all obstacles are represented as a force. The stronger the force, the bigger the obstacle is. For navigation, the robot just tries to minimize the resulting force on him so he can drive safely around obstacles with an optimal distance to each obstacle.

Occupancy grids The map is divided into cells and each cell is either labelled as drivable or not drivable. This approach is mostly used in map-based and map-building systems. Then the robot can perform a path planning algorithm to find its way through the environment.

Stereo 3D reconstruction Stereo 3D reconstruction is based on two cameras which are physically distant. They work much like the human vision. Out of the information of the distance of the two image feeds and the actual differences in the two image feeds one can calculate an exact 3D model of the surrounding. Based on this 3D representation the navigation can be handled. This approach is wildly used in almost all systems because it is extremely cheap and powerful. Figure 2 shows such an algorithm in action.

Optical Flow Optical Flow is based on tracking the movement of brightness patterns. With this information and the information about the distance to certain patterns (via 3D reconstruction or something similar), the exact speed relative to the robots surrounding can be calculated. This approach is also wildly used in all systems and is even sometimes combined with other systems like stereo 3D reconstruction.

Appearance based In the appearance based approach, the robot is provided with images of certain obstacles which should be avoided. For example the robot may know an image of a chair. As soon as it recognizes something in its image feed which is close to the chair image, it knows that this area in front of it represents an obstacle.

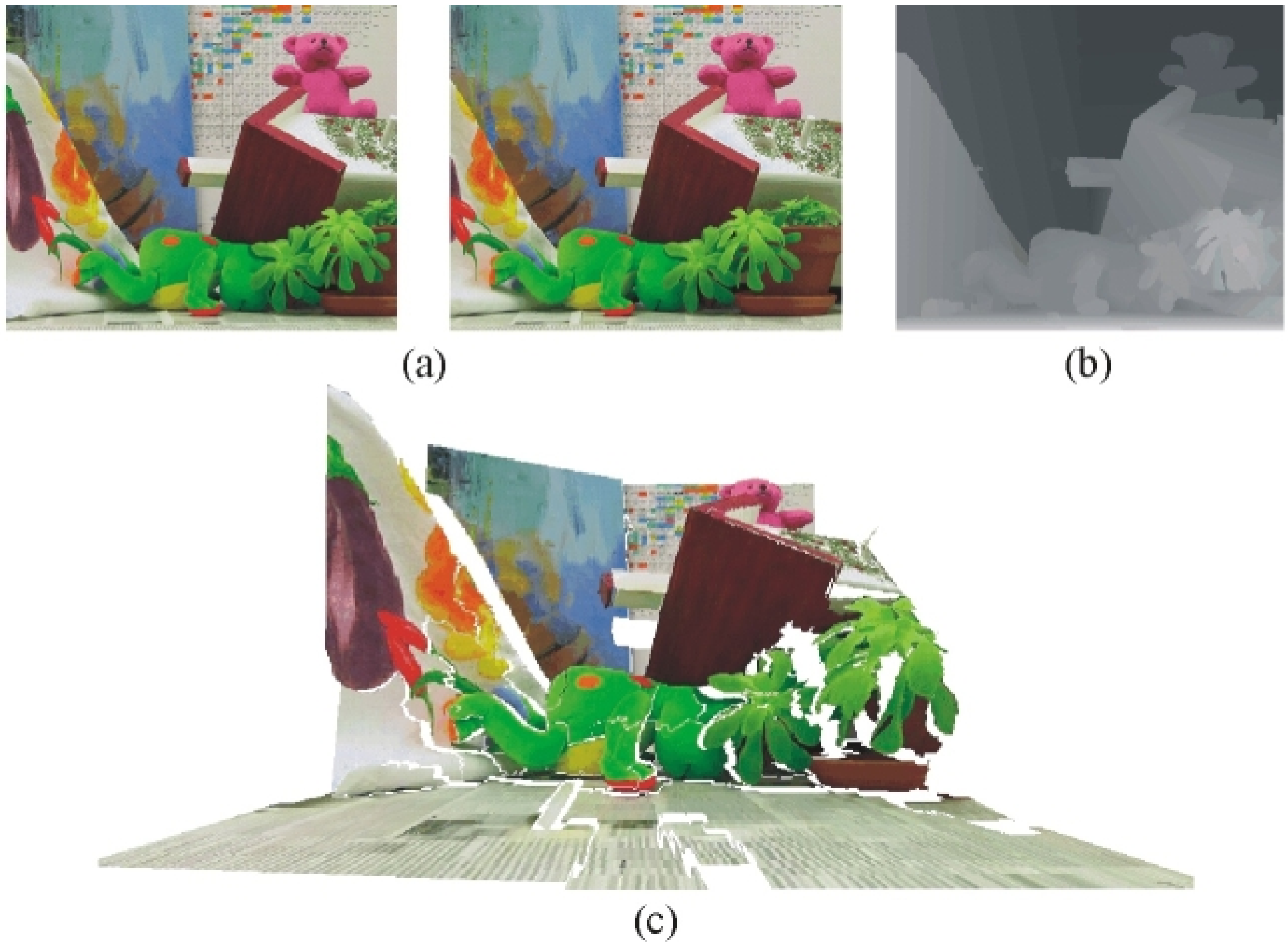


Figure 2: The principle of 3D reconstruction via stereo vision. (a) Left and right views of a stereo pair. Due to the different viewpoints, corresponding points in the two images are displaced in horizontal direction. (b) Amount of horizontal pixel displacement between the input views. (Large displacements are represented by bright pixels.) The amount of displacement is inversely proportional to a pixel's depth in the scene. The image of (b) is therefore sufficient for generating the 3D scene reconstruction of (c).[9][11]

Problems in indoor approaches When considering indoor approaches, we run into a certain problem which needs to be handled by all system. Indoor systems cannot rely on GPS so it is very hard for these systems to estimate their speed and direction. One can propose a system which measures the speed of every wheel but this is not sufficient due to multiple reasons. First, one cannot exactly measure the speed of every wheel. This effect is shown in figure 3. Another problem which normally does not arise in indoor situations but especially in space where we also do not have GPS is wheel slipping. So it might be that one wheel slips because of sand and therefore delivers incorrect movement data.

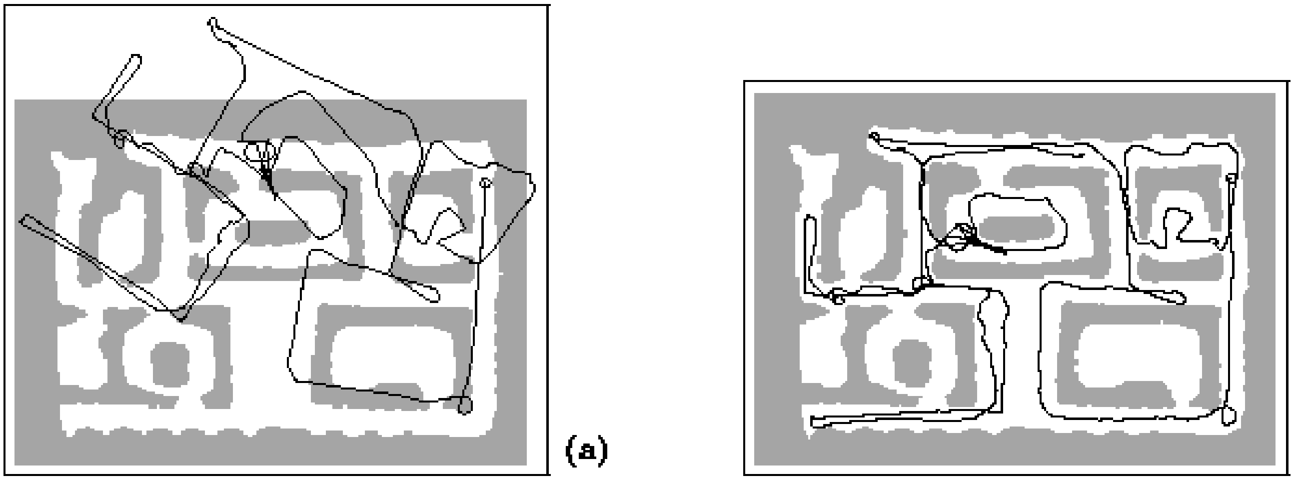
NAVIGATION IN SPACE

System which operate in space have extreme constraints regarding their reliability and their weight. Since an additional kilogram of equipment can cost up to 40'000$, one goal of these system is to be extremely light. But since such a system is not helpful at all if it stops working after a year, reliability is graded more important than weight. These systems such as the Curiosity robot[4] rely mostly on 3D reconstruiction of the terrain surrounding the robot. In the case of Curiosity, the robot is equipped with eight so called Hazcams[5]. These cameras are mounted in pairs of 2 all around the robot and provide it with low resolution black and white images. The reason behind mounting pairs of two is nothing more than reliability. If one camera breaks, the other one can be used. Furthermore the Curiosity robot contains two pairs Navcam[6] which take 360 degree panoramic images and are used for path planning, using again the stereo 3D reconstruction algorithm.

UNDERWATER NAVIGATION

When considering systems which operate under water, we have established systems which are equipped with SONAR[7]. But these systems run into problems when autonomous oper-

[4] A Mars rover mission started in 2012 with a multi-billion dollar budget
[5] short for hazard avoidance cameras
[6] Navigational camera
[7] Sound Navigation And Ranging



(a) (b)

Figure 3: In image (a) is the path shown how the robot perceived it just by the movement of the wheels and steering commands. In image (b) is the path shown that the robot actually drove[12]

ation is required for a long time. Since vision based system can be very cheap and very power efficient, theses systems are favoured. But using vision for navigation purposes can be a difficult task under water. Sometimes, there are almost no reference point which can be used to track the robots movement. Underwater currents even increase the difficulty of this task since the robot moves but his whole surrounding moves with the robot.

AUTONOMOUS DRIVING

When talking about autonomous driving there are certain requirements for such a system. These requirements range from reliable pedestrian detection, obstacle detection over road sign detection to street signal detection. These are absolutely necessary for safe operation of the robot. So in the next sections the most used approaches for each of these requirements are presented.

Reliable pedestrian detection

When talking about pedestrian detection the most important aspect is, that every pedestrian is detected. We have seen multiple algorithms which suit for such a purpose such as stereo 3D reconstruction. But as seen in figure 4 this gets pretty difficult, due to the sheer amount of pedestrians. As mentioned the most common techniques use stereo 3D reconstruction[5], but there exist also algorithms which are based on shape recognition[7]. Another very interesting approach, mentioned in [6], is predicting the pedestrian motion. In this paper the author proposed a system which predicts the pedestrians motion to increase the safety of the system.

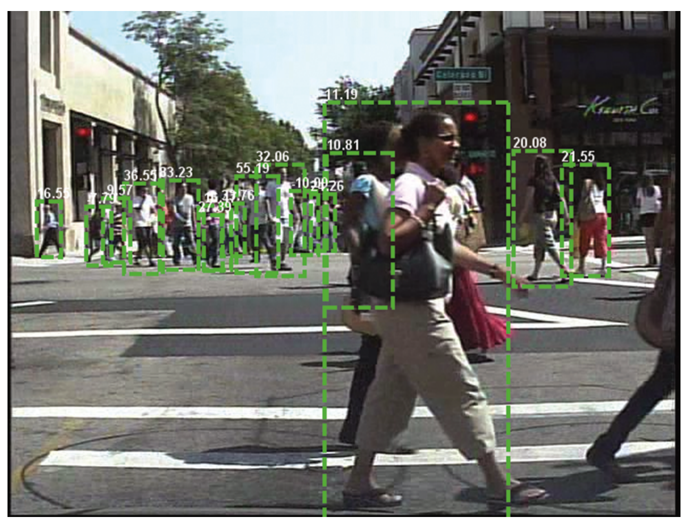


Figure 4: An example of how difficult it can be to detect all pedestrians[13]

Detect and interpret road signs

Information about the current speed limits are normally stored in the maps provided to the robot. These maps are not hundred percent accurate so they need to be improved by a visual system which detects the actual location of such a sign. There are 2 approaches to detect signs which are presented in [7]. They use stereo 3D reconstruction to search for signs

Figure 5: An example showing traffic light detection[14]

and detection based on shape, color and motion. The second approach is just a simple template matching on the input image.

Obstacle detection

Even when driving on a pavement road and detecting all pedestrians, there can still be certain obstacles which need to be detected such as speed bumps, trees on the street and so on. Especially the speed bump detection is a rather difficult task. A normal speed bump is only about 20 centimetres high so when for example building an obstacle map, one must be very careful when thresholding.

Road following

Road following is manly done by detecting dark-light-dark transitions. These correspond to the lines which border the lane, so by detecting these, the robot knows exactly what the direction of the street is. This information is manly used to improve the information provided by the map. This algorithm is presented in more detail in [5]

Street signals

When reacting to street signals one must be sure to react to the correct signals. Figure 5 shows a system which is able to detect all street lights which affect the line the robot is driving on. This algorithm was presented in [8].

CONCLUSION

As one can see there are pretty reliable systems out there which are, in addition, also extremely cheap due to their usage of vision based systems. Future research will hopefully increase the amount and the performance of vision based systems. As mentioned these systems are extremely cheap so to be competitive in the consumer market, systems must be developed which are cheaper than using lidar or radar based systems.

ACKNOWLEDGMENTS

I want to thank Marian George who acted as my supervisor for her great assistance and help in creating the presentation. She provided me with clarification when something was unclear and provided good advice on creating the presentation.

REFERENCES

1. Francisco Bonin-Font, Alberto Ortiz, Gabriel Oliver, Visual Navigation for Mobile Robots: A Survey. In *Journal of Intelligent and Robotic Systems '08* November 2008, Volume 53, Issue 3, pp 263-296.

2. Larry Matthies, Mark Maimone, Andrew Johnson, Yang Cheng, Reg Willson, Carlos Villalpando, Steve Goldberg, Andres Huertas, Andrew Stein, Anelia Angelova, Computer Vision on Mars. In *JInternational Journal of Computer Vision '07* October 2007, Volume 75, Issue 1, pp 67-92.

3. Jonathan Horgan and Daniel Toal, Computer Vision Applications in the Navigation of Unmanned Underwater Vehicles. In *Underwater Vehicles* Book edited by: Alexander V. Inzartsev, ISBN 978-953-7619-49-7, pp. 582, December 2008, I-Tech, Vienna, Austria.

4. Sebastian Thrun, Sven Strohband, Gary Bradski, Pamela Mahoney et al., Stanley: The robot that won the DARPA Grand Challenge. In *Journal of Robotic Systems '06* Special Issue on the DARPA Grand Challenge, Part 2 archive Volume 23 Issue 9, Pages 661-692, September 2006.

5. Jaewoong Choi, Junyoung Lee, Dongwook Kim, Giacomo Soprani, Pietro Cerri, Member, IEEE, Alberto Broggi, Senior Member, IEEE, and Kyongsu Y, Environment-Detection-and-Mapping Algorithm for Autonomous Driving in Rural or Off-Road Environment. In *IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS '12* VOL. 13, NO. 2, JUNE 2012

6. Christian Berger, Bernhard Rumpe, Autonomous Driving-5 Years after the Urban Challenge: The Anticipatory Vehicle as a Cyber-Physical-System. In *Proceedings of the 10th Workshop on Automotive Software Engineering '12* pp. 789-798, September 2012.

7. Uwe Franke, Dariu Gavrila, Steffen Gorzig, Frank Lindner, Frank Paetzslld, and Christian Wohler, Autonomous Driving Goes Downtown. 1998

8. Jesse Levinson, Jake Askeland, Jan Becker, Jennifer Dolson, David Held, Soeren Kammel, J. Zico Kolter, Dirk Langer, Oliver Pink, Vaughan Pratt, Michael Sokolsky, Ganymed Stanek, David Stavens, Alex Teichman, Moritz Werling, and Sebastian Thrun, Towards Fully Autonomous Driving: Systems and Algorithms. In *IEEE Intelligent Vehicles Symposium '11* Baden-Baden, Germany, June 5-9, 2011.

9. Margrit Gelautz, Michael Bleyer, Liu He, Nicole Brosch, Evaluation and Design of Energy Functions for Global Stereo Matching. 2011

10. Taken from [4]

11. Taken from [9]

12. http://www.cs.cmu.edu/

13. http://opticalengineering.spiedigitallibrary.org/article.aspx?articleid=1

14. Taken from [8]