18 Computer Vision Applications in Service Robots

Dimitrios Koikas
Dpt. of Automation Engineering, Piraeus University of Applied Sciences, Aigaleo, Greece, E-mail: dpkoikas@gmail.com

Grigoris Nikolaou
Dpt. of Automation Engineering, Piraeus University of Applied Sciences, Aigaleo, Greece, E-mail: nikolaou@puas.com

18.1 Abstract

Over the past few years, service robots development has greatly advanced. An important aspect is the visual perception of the robot's environment. Hence, research on service robots is focused on computer vision algorithms and techniques. This paper aims to present and analyse the various approaches on the forementioned direction. Specifically, this paper focuses on people tracking and identification techniques. Data for computer vision algorithms are collected by sensors such as cameras (RGB, Depth etc.) and laser scanners. These techniques vary depending on the kind of data we acquired and of course, our aim in each case. Machine learning is the tool most frequently used, along with image processing and comparison.
Keywords: computer vision, people tracking/identification, service robots

18.2 Introduction

By principle, service robots are used to help the elderly in their home environment. Of course, this does not exclude the possibility to generally serve people in their homes. It is obvious then, that robots are required to detect, identify and track persons. Identification and tracking are critical for the best possible execution of any task. Their significance is better understood if real life scenarios are taken into consideration. For example, the robot must be able to identify which person is instructing it at any given time. Thus, the robot could either execute or ignore the command. Another important aspect is security. In case there is an intruder in the environment (or trying to intrude), the robot should be alarmed and execute an emergency plan, in order to prevent malicious situations.

18.3 Sensors

There is a huge variety of sensors utilised to perform vision related tasks. Of course, cameras are the heart of these systems. The point is to choose an appropriate one, that can provide the developer with data whose input will make sense and attribute to the algorithmical approach he wants to use.
Cameras are usually sorted by the nature of the image in their output. Mostly known types are: a) grayscale, b) Red-Green-Blue (RGB) and c) RGB-Depth. In service robots, RGB-D cameras are more commonly encountered. These sensors offer the flexibility to have various data types available and use them as anyone may, in different tasks running simultaneously. Microsoft Kinect (both versions) and Asus Xtion Pro are the most popular.

***Figure 1**: Most popular RGB-D sensors [F1][F2]*

Another useful sensor is the laser scanner. It is primarily used for environment mapping and navigation. However, it can be used for people detection and tracking too. For this purpose, pattern recognition techniques are trying to learn and recognize the pointclouds belonging to human legs. Again, a variety of data is obtained via the sensor, such as position in 3D-space and intensity.

***Figure 2**: Indicative laser scanners [F3][F4]*

18.4 Mapping and Navigation

In a domestic environment, a robot must have the ability to navigate without damaging the surroundings and, most importantly, without hurting any people. After accomplishing this, the robot is able to follow persons around and navigate itself to any position that is called or needed.
The problem described above is called Simultaneous Localization and Mapping (SLAM). The challenge is to construct and update a map of the environment, while, as the name implies, keeping track of the robot location. By robot location, both the position and orientation of the robot are described. Many SLAM algorithms have been developed for various scenarios and applications, such as Unmanned Aerial Vehicles (UAVs) and Service Robots. Moreover, not all algorithms use the same sensors. So, deciding which algorithm to put in use is affected by the equipment available and vice versa.
As formentioned, laser scanners are important tools for these problems. The idea behind its use is quite simple. As the laser scans the environment, the laser beams are reflected back to it when an obstacle is met. The distance is measured using the Time of Flight Principle. The conclusion is that this distance is an empty space, so the robot considers it as this. If the obstacle is moved, the map is updated with the new conditions.

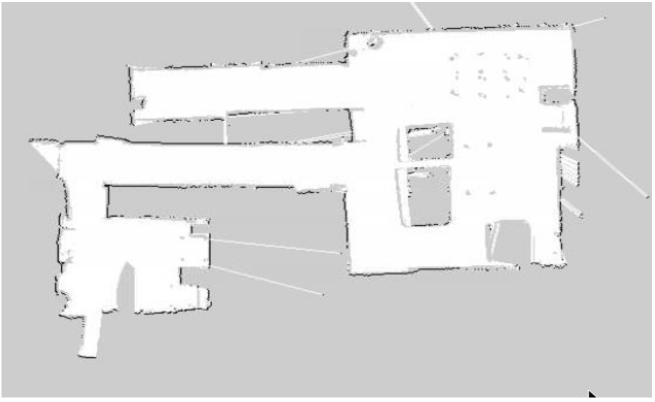

***Figure 3**: Example SLAM algorithm output map [F5]*

18.5 <u>People detection and identification</u>

The process of detecting people is crucial for serving the person in need. The robot has to be aware of the person's position in order to bring items to him. Moreover, it needs identify the target person, in case there are other people in the environment, for any reason.
Detecting people can be accomplished using pattern recognition with a laser scanner, as mentioned in section 2. There are more effective ways of doing so nevertheless. Some techniques are trying to detect the human body "piece-by-piece". It means that the algorithm is looking for a head, a torso, two arms and two legs. To accomplish these detections, usually Machine Learning-trained classifiers are used. This can be applied to directly detecting the whole human body too.
Face recognition is another robust approach. The Haar-Cascade algorithm is the most popular way to train classifiers for this task. The robot is expected to recognize the face of the operator among others and deliver to him. Face/head detection (and recognition in case of face) is also used to cross evaluate the data for people detection. For example, sometimes legs can be detected from a laser scanner. But, if no head is detected above them, then the assumption of person existence is false. Thus, the false positive leg detection is filtered out of the system.
Picture characteristics comparison is interesting too. When knowing the visual characteristics of a person and his identity is given, it is possible to compare these with any other person to re-identify him. When a person enters the robot's field of view, the comparison process begins. If they match in a large degree, then the "new entry" is actually an already known person. Again, this can become more robust using face recognition algorithms.

18.6 <u>Object detection and recognition</u>

Object perception is an important aspect for most service robots. The concept of perception includes both the position and object orientation. Orientation knowledge is crucial when it comes to manipulating the object. Manipulation is usually executed with a robotic arm gripper. So, the robot must be able to decide how to grip the object, in case of shape abnormalities or if special handling is required. Object perception algorithms vary, just as people identification case. Again, the use of algorithms is affected by the equipment available.
At this point, it should be mentioned that the Haar-Cascade Classifier [see P1] is able to be used in this kind of tasks as well.
A simple approach is detecting edges in images. Edge detection is based on brightness values. Discontinuity in brightness suggests that an edge exists in this line segment. Only sharp changes are taken into account when it comes to discontinuity. Then, the edges are connected and the contour of the object takes shape. The most popular algorithm is the Canny Edge Detector [see P2].

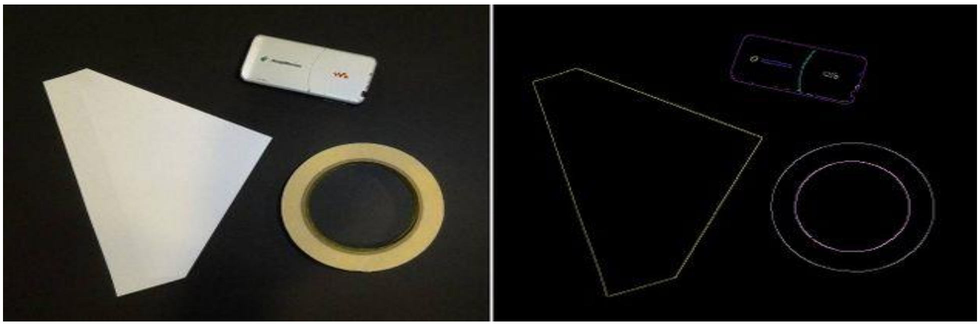

Figure 4: *Contour detection [F6]*

Another algorithm is based on plane detection. At first, surface planes are detected. Then, point clusters above this plane are considered as objects, after cross-matching some criteria of course. The drawback is that objects with planar shape (e.g. books, calculators) are hard to detect, especially when they are thin. Of course, the viewing angle affects the results. If the plane and camera are placed in parallel, there is a high chance that many points are not detected.
Scale Invariant Feature Transform (SIFT) [see P9] extracts features from an image. This way, it creates a description consisted of these features corresponding to the training sample. If the features extracted are not affected by image scaling and noise of any kind, then the recognition is more reliable. Also, feature position is of great importance. Relative positions between the training sample and the image in comparison should not be different or diverge a lot.
Speeded Up Robust Features (SURF) [see P10] is a feature extractor too. The difference is that it outperforms SIFT when it comes to computation speed and image transformation robustness. Roughly explained, the algorithm follows three steps: a) point detection, b) point neighbourhood detection and c) feature matching between training sample and comparison image.

18.7 Conclusions

Computer vision is a constantly growing field that offers solutions in various ways. It is interesting that for any given problem, there can exist more than one approaches. This makes it a fascinating subject, pushing the limits of creativity and imagination even further. The fact that research is focused on developing more and more reliable systems, gives confidence for a future algorithm or solution that solves multiple problems at once.

18.8 References

[P1] Paul Viola, Michael Jones, “Rapid Object Detection using a Boosted Cascade of Simple Features“, *Accepted Conference on Computer Vision and Pattern Recognition*, 2001

[P2] John Canny, “A Computational Approach to Edge Detection”, *IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. PAMI-8, No.6, November 1986*

[P3] William Robson Schwartz, Aniruddha Kembhavi, David Harwood, Larry S. Davis, “Human Detection Using Partial Least Squares Analysis”, ICCV, 2009

[P4] Paul Scovanner, Marshall F. Tappen “Learning Pedestrian Dynamics from the Real World”, ICCV, 2009

[P5] Christoph H. Lampert, “Detecting Objects in Large Image Collections and Videos by Efficient Subimage Retrieval”, ICCV, 2009

[P6] Gang Hua, Amir Akbarzadeh, “A Robust Elastic and Partial Matching Metric for Face Recognition”, ICCV, 2009

[P7] Jean Ponce, Martial Hebert, “On Image Contours of Projective Shapes”, ECCV, 2014

[P8] Feng Zhou, Jonathan Brandt, Zhe Lin, “Exemplar-based Graph Matching for Robust Facial Landmark Localization“, ICCV, 2013

[P9] David G. Lowe, “Distinctive Image Features from Scale-Invariant Keypoints”, International Journal of Computer Vision, 2004

[P10] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool, “SURF: Speeded Up Robust Features”

[F1] https://images-na.ssl-images-amazon.com/images/I/31raUar9A4L._SX466_.jpg

[F2] http://img.game.co.uk/ml2/3/3/7/2/337275_scr1_a.png

[F3] http://www.mrpt.org/wp-content/uploads/2013/10/SICK_LMS100.jpg

[F4] https://www.hokuyo-aut.jp/02sensor/07scanner/img/utm_30lx_top.jpg

[F5] http://www.ctralie.com/Teaching/DukeDusty2/apartmentmap.jpeg

[F6] http://boofcv.org/images/thumb/9/99/Object_contours.jpg/631px-Object_contours.jpg