Depth camera driven mobile robot for human localization and following

Nikolaos Skordilis*, Nikolaos Vidakis* Georgios Triantafyllidis†, Lazaros Nalpantidis‡
*Department of Applied Informatics and Multimedia, Technological Educational Institute of Crete, Greece
†Department of Architecture, Design and Media Technology, Aalborg University Copenhagen, Denmark
gt@create.aau.dk
‡Department of Mechanical and Manufacturing Engineering, Aalborg University Copenhagen, Denmark
lanalpa@m-tech.aau.dk

***Abstract*—In this paper the design and the development of a mobile robot able to locate and then follow a human target is described. Both the integration of the required mechatronics components and the development of appropriate software are covered. The main sensor of the developed mobile robot is an RGB-D camera. This sensor is usually combined with the appropriate middleware that can detect humans in a scene and also provide the position of the detected human in the 3D space. One of the cues this middleware's algorithms are using to detect humans is motion, thus resulting in many false detections when applied to data captured by a mobile platform. This work proposes the use of a special-tailored feed forward neural network to further process the initial detections, identifying and rejecting most false positives. Experimental results based on two self-captured data sets show the improved detection rate of our two-stage human detector, as well as the efficient real-time performance of the proposed system for human localization and following.**

Fig. 1. The developed robot following a human

I. Introduction

While robot technology is rapidly advancing, robots are in most of the cases used in well-specified applications and restricted in confined places, e.g. in industrial plants or labs. Intelligent robots for general domestic use are still not mature enough. The robots that will be allowed to operate in our homes and offices are human-friendly robots, able to safely coexist with us and support us effectively [1]. One of the most important aspects in the development of human-friendly robots is to achieve cooperation between humans and robots [2], [3]. This symbiosis presumes that robots will be able to reliably detect and robustly follow humans. Another reason that makes reliable detection of humans so important is safety [4], [5]. Upon the detection of humans robots should adapt their operation, e.g. lowering their speed or replacing their paths, so as to ensure the safety of people [6].

While the use of simple cameras could keep the cost of the developed robot low, the sole use of 2D images can cause many problems in robotic applications as the ones considered in this work. The luminance of the area, shadows, uncertainty about the structure of the scene are just some of the issues that produce extra trouble to an object and human detection and recognition module that uses simple camera input. Also, the image-based extraction of the object distances was a difficult problem to be solved with 2D image processing techniques; the object distances are vital for estimating scale, which is necessary for the object recognition. All these problems can be easily solved with the use of accurate depth maps. Laser scanners that can provide depth maps are expensive, but now with the availability of RGB-D sensors such as the MS Kinect or the Asus Xtion Pro there is a cheap and efficient way to get a depth map [7].

This work describes the development of a self-made mobile robot equipped with an RGB-D sensor, i.e. an ASUS Xtion Pro. Using the 3D imaging capabilities of this sensor, the robot is able to locate a person in the scene and then follow him. PrimeSense's OpenNI middleware library can be used for detection and tracking of humans, providing their relative position in the 3D space [8]. This library is today a standard choice for such applications. However, its use is not always free of problems. Many false results of the OpenNI's human detector and tracker are due to the fact that motion is (among others) a cue for determining a human's existence. Thus, if an object seems to move, it is often erroneously detected as a human. While this assumption might be reasonable in a home entertainment system, where the sensor is firmly positioned, it produces many problems in a mobile system. In this latter case, the whole scene seems as if moving due to the motion of the robotic platform, and as a consequence many false detections occur.

Furthermore, in this work we explore a solution to the false human detections problem, that involves the additional use of a specially tailored feed-forward Neural Network (NN). To efficiently handle the continuous stream of image and depth data, only some representative features are kept, necessary for the detection of humans. Using these features as the input of the NN, the results of the OpenNI tracker are validated and verified. Finally, upon detection of a human, the robot automatically moves so as to keep the human in the center of its horizontal field of view and within a desired distance.

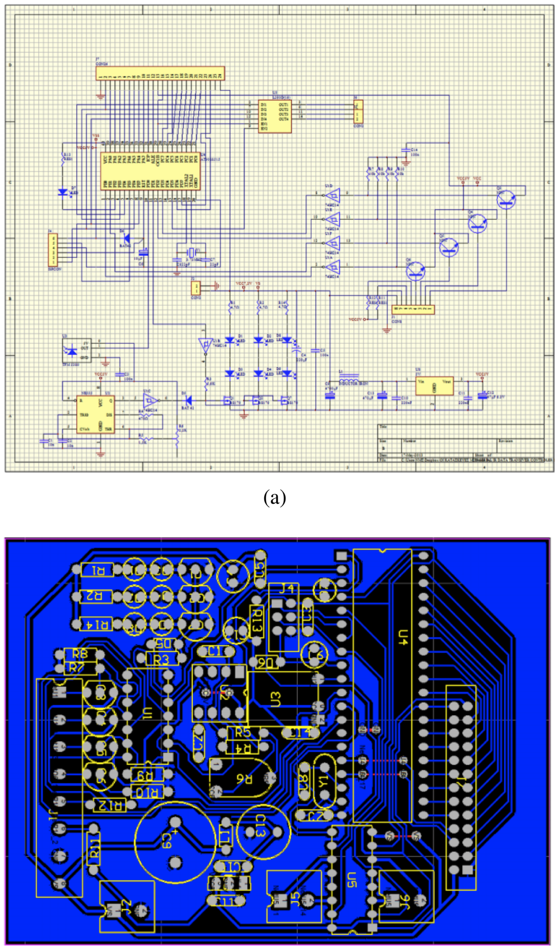
A. Related Works

In this context, there are many applications in robotics that need mobile robots that are capable of locating or/and following humans. For example, the robotic assistant in [9] need to be close to and aware of the position of the person who commands it. In some other cases the ability to follow a human is enough, for example carrying objects [10]. Also, such a behavior can be used in games with robot, e.g. in a robotic pet [11]. Several research groups have presented work on robots designed to follow humans for over a decade. The work of [3] describe a mobile robot, which always faces humans and acts as an assistant robot. The mobile robot presented in [9] is a human-following robot for assisting humans. This robot is aimed at guiding a wheelchair in a hospital or a station and has the ability to estimate the position and velocity of humans and to avoid obstacles. In [11], four-legged mobile robots for following humans are considered. However, most of these studies only addressed the problem of how to follow humans once detected, not how to naturally detect the presence of humans at first hand. In general, various approaches have been proposed to detect humans, as well as the relative position between a mobile robot and a human. The use of wearable markers for the detection and distance estimation of humans has been reported [12], but such solutions require a lot of preparation and are thus not natural. The sole use of sensors, without any prerequisites from the human subjects is a preferable direction. Such methods may exploit ultrasonic sensors, auditory sensors, laser range sensors, infrared cameras, visual cameras, and so on [13]. Gockley et al. [14] discuss a laser-based person-tracking method and two different approaches to person-following: direction-following and path-following.

II. Hardware Integration

For the scope of this work, a self-made mobile robot was developed. The adopted design for the robot consists of a notebook, a electrically powered toy vehicle, and a circuit that is controlled from the notebook via serial port and drives the motors of the vehicle. Also, a USB to serial converter is used on the robot, as well as an ASUS Xtion Pro RGB-D sensor. The core of the control circuit is an AT90S8515 AVR microcontroller. A special PCB, seen in Fig. 2 was developed to host the micro controller and all the required electronics that drive the vehicle motors. The notebook is a 10” one, equipped with 2GB RAM and an Intel Atom N570 processor operating at 1.66 GHz.

The total weight of the mobile robot with all the batteries, circuitry and sensors is around 4 Kgr. The track length is 32 cm, the total length of the mobile robot is 41 cm, the height up to the notebook is 20 cm and when also considering the Xtion sensor the height is 28 cm. The width of the platform is 22 cm. Fig. 3 shows two photos of the developed mobile robot.



(a)

(b)

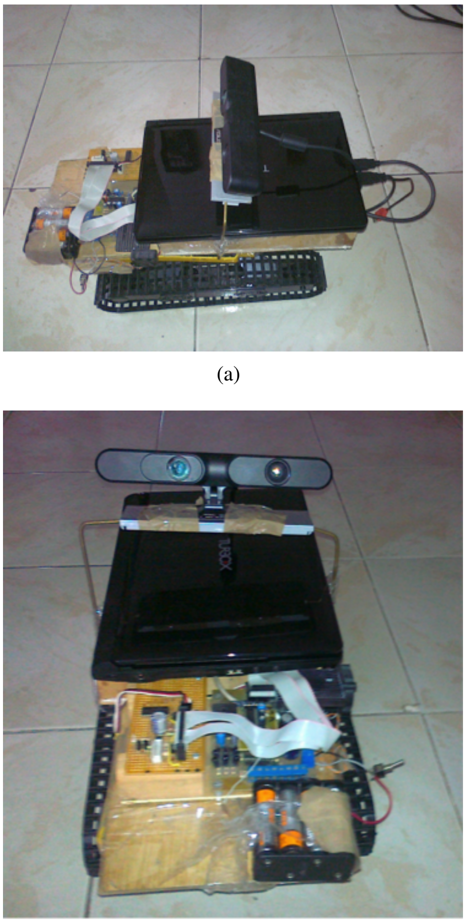
Fig. 2. Schematic and development of PCB

III. Software Development

A. Human Detection and Neural Network

To detect humans using a depth camera we employ the standard the commonly used, free, cross-platform driver OpenNI and the NiTE human tracking library [8]. These functions provide us with the coordinates of the human, a depth map of the scene and an image that has the pixels of the depth map belonging to the human highlighted. Using all these kind of the data and information as inputs to our neural network is not efficient. To discern between real humans and objects falsely detected as humans (mainly due to the sensor ego-motion), one can safely first get rid of much redundant information contained in the input stream. As a result, a process is required that will decrease the amount of data, keeping only the necessary and meaningful for the required human detection. A desired attribute of such a process would be to provide data that do not significantly depend on the human’s pose or position.

Towards this goal, we count the number of pixels per image line belonging to the object detected as human from the OpenNI detector. This number (which is actually the object’s width) is not significantly dependent on the object’s position and motion along the in x axis. Then, to further



(a)

(b)

Fig. 3. Views of the integrated self-made robot

decrease the amount of data, sub-sampling is used to gather this information. Only 16 lines are examined and only 16 pixels per line. However, experimentation has shown us that just this is not enough to recognize humans effectively and, as a consequence, one more feature had to be added as input in the neural network. Each of the 16 lines is scanned and the number of transitions from pixels belonging to the object to pixels not belonging to it, or vice versa, is counted. This feature don't change for small changes of the object position along the x and z axes.

Finally, the inputs of the considered NN were determined to be:

- the 3D position of the object centroid, as observed by the RGB-D sensor,
- the number of pixels belonging to the object per line of the sub-sampled image,
- the number of detected transitions per line of the sub-sampled image.

The suggested neural network has 35 neurons at the input layer, 85 neurons at the hidden layer and 2 at the output layer. The neurons transfer function is tansig. The one output means human and the other output not.

B. Robot Operation

This section describes how the mobile robot is behaving when operating. The robot is constantly trying to detect humans. If it can't detect any human for a set amount of time, it starts rotating slowly with in hope that a human is somewhere out of the field of view.

After the detection of a human we use the coordinates of the detected human to move the mobile robot accordingly. Please notice that the image used in our experiments is of QVGA resolution 320x240. So, if the x coordinate is greater than 200 the mobile robot rotates right in order to observe the human close to the center of its 320 pixel wide frame, else if x is less than 120 it rotates left. If x is within the range [120, 200], then if the human is close enough (i.e. in our experiments z¡1000) the mobile robot stops. Else if z is less than 1400 the speed of the motors is lowered proportionally to the difference of z from 1000. If z is larger than1400 maximum speed is chosen. If x is within the range [120, 200] the balance of the speed between left and right motor is determined as follows: If x=160 the speed is equal, if x is greater than 160 left motor is given more power, according to the difference. If x is less than 160 the right motor takes more power using to the same rule. The proposed scheme is illustrated in Fig. 4.

IV. Experimental Evaluation

As part of our experimental procedure we used two sets of samples for training and testing the neural network. At first 775 samples (depth maps) were used with a detected human as provided from the OpenNI library (either true or false human detection). Each sample contains the central position of the detected object (human), its width and the number of changes from object pixels to non-object pixels or the the other way around. In 259 of these samples there is a human in the scene, while in the rest 516 samples there is a false detection of a human. The second set of samples used for training and testing contains 4196 samples, out of which 2909 samples present a true human detection and 3425 samples present a false human detection. Table I summarizes the two gathered data sets.

TABLE I. Detection results of the initial OpenNI human detector

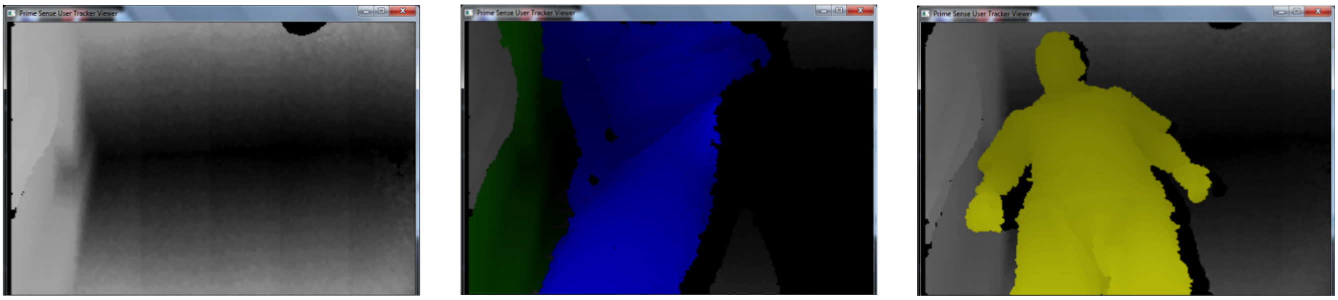
Dataset	# of samples	True Positives	False Positives
Set A	775	256	516
Set B	4196	2909	3425

Some of the gathered data are shown in Fig. 5, as captured during the operation of the proposed system.

Table II illustrates the obtained percentage of false human acceptance and rejection after applying the proposed neural network. During our experiments, 70% of the samples were used for training, 15% for validation and 15% for testing. When using the set B, the performance of the proposed neural network is very satisfactory.

Overall, the proposed mobile robot design provides efficient detection and following of a human at maximum distance of about 4 meters. A video showing the mobile robot following a human can be found online[1].

[1] http://www.youtube.com/watch?v=-lTHV6kswWA



(a) True negative (b) False negative (c) True positive

Fig. 5. Human detection results

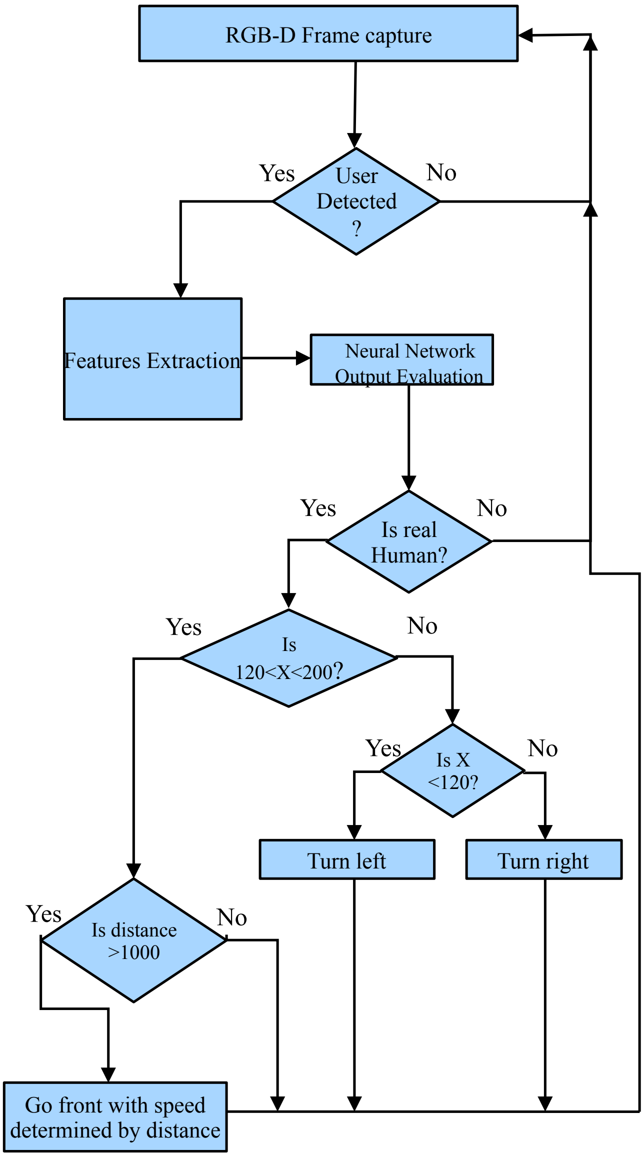


Fig. 4. Flow Chart

TABLE II. DETECTION RATE FOR THE PROPOSED ALGORITHM

Dataset	# of samples	False Human Acceptance	False Human Rejection
Set A	775	4.7%	4.0%
Set B	4196	2.6%	1.4%

V. CONCLUSION

In this work we have presented the development of a little mobile robot that is looking for the presence of humans and the follows them. The problem of frequent false human detections, when employing RGB-D sensors on mobile platforms, was addressed by developing and using a suitable neural network. For efficiency purposes, only selected features form the continuous stream of available data were fed as inputs to the neural network. The result of the proposed system was that false initial human detections were in most of the times filtered out, while very little additional false rejections were added. The promising results and the nice real-time operation of our mobile robot shows that simple systems and solutions like the ones described in this work, can be used as basic building blocks towards the ultimate goal of safe human-robot symbiosis.

REFERENCES

[1] R. Bodor, B. Jackson, and N. Papanikolopoulos, "Vision-based human tracking and activity recognition," in *Mediterranean Conference on Control and Automation*, 2003, pp. 18–20.

[2] H. Sidenbladh, D. Kragic, and H. I. Christensen, "A person following behavior for a mobile robot," in *IEEE Iternational Conference on Robotics and Automation*, 1999, pp. 670–675.

[3] S.-O. Lee, M. Hwang-Bo, B.-J. You, S.-R. Oh, and Y.-J. Cho, "Vision based mobile robot control for target tracking," in *IFAC Workshop on Mobile Robot Technology*, 2001, pp. 73–78.

[4] T. Kabe, K. Tanaka, H. Ikeda, and N. Sugimoto, "Consideration on safety for emerging technology – case studies of seven service robots," *Safety Science*, vol. 48, no. 3, pp. 296–301, 2010.

[5] E. Mitka, A. Gasteratos, N. Kyriakoulis, and S. G. Mouroutsos, "Safety certification requirements for domestic robots," *Safety Science*, vol. 50, no. 9, pp. 1888–1897, 2012.

[6] J. L. Crowley, "Planning and execution of tasks for a domestic robot," *Robotics and Autonomous Systems*, vol. 5, no. 3, pp. 257–272, 1989.

[7] F. Hoshino and K. Morioka, "Human following robot based on control of particle distribution with integrated range sensors," in *IEEE/SICE International Symposium on System Integration*, 2011, pp. 212–217.

[8] "NITE - natural interface technology for end-user," http://www.primesense.com/solutions/nite-middleware/.

[9] E. Prassler, D. Bank, B. Kluge, and M. Hagele, "Key technologies in robot assistants: Motion coordination between a human and a mobile robot," in *International Symposium on Robotics*, 2001, pp. 410–415.

[10] J. Stückler and S. Behnke, "Following human guidance to cooperatively carry a large object," in *IEEE-RAS International Conference on Humanoid Robots (Humanoids 2011)*, Bled, Slovenia, 2011, pp. 218–223.

[11] M. Fujita, "Aibo: Toward the era of digital creatures," in *International Journal of Robotics Research*, vol. 20, no. 10, 2001, pp. 781–794.

[12] M. Bakar, R. Nagarajan, and A. Saad, "Development of a doctor following mobile robot with mono-vision based marker detection," in *IEEE Applied Power Electronics Colloquium (IAPEC)*, 2011, pp. 86–91.

[13] Q. K. Dang and Y. S. Suh, "Human-following robot using infrared camera," in *International Conference on Control Automation and Systems*, 2011, pp. 1054–1058.

[14] R. Gockley, J. Forlizzi, and R. Simmons, "Natural person-following behavior for social robots," in *ACM/IEEE International Conference on Human-Robot Interaction*, 2007.