Development of a visually guided mobile robot for environmental observation as an aid for outdoor crisis management operations

Geert DE CUBBER, Daniela DOROFTEI and Gabor MARTON
Department of Mechanics, Royal Military Academy of Belgium;
Av. De La Renaissance 30, 1000 Brussels, Belgium;
geert.de.cubber@rma.ac.be, daniela.doroftei@rma.ac.be

Introduction

Crisis management teams (e.g. fire & rescue services, anti-terrorist units, ..) are often confronted with dramatic situations where critical decisions have to be made within hard time constraints. A complete overview of the crisis site is necessary to take correct decisions in these circumstances. However, obtaining such a complete overview of a complex site is not possible in real-life situations when the crisis management teams are confronted with large and complex unknown incident sites. In these situations, the crisis management teams typically concentrate their effort on a primary incident location (e.g. a building on fire, a wreckage, ...) and only after some time (depending on the manpower and the severity of the incident), they turn their attention towards the larger surroundings, e.g. searching for victims scattered around the incident site. A mobile robotic agent could aid in these circumstances, gaining valuable time by monitoring the area around the primary incident site while the crisis management teams perform their work. However, as the human crisis management teams are in general already overloaded with work and information in any medium or large scale crisis situation, it is essential that such a robotic agent - to be useful - does not require extensive human control (hence it should be semi-autonomous) and it should only report critical information back to the crisis management control center. In this paper, we discuss the development of such a semi-autonomous outdoor mobile robot, which is able to search for human victims on an incident site, while navigating semi autonomously, using stereo vision as a main source of sensor information. The design and development of such a robotic agent raises 2 main questions:

1. How can we detect human victims lying unconscious or partly conscious on the ground, only using visual information from a stereo camera system, and this in dynamic outdoor illumination conditions?

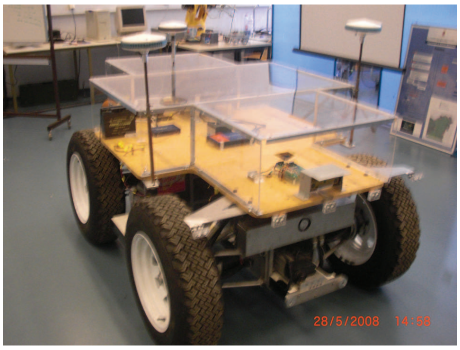

Figure 1. The RobuDem platform with GPS and stereo vision system, used for evaluating the presented algorithms.

2. How can the robot be made semi-autonomous, such that it can handle a high-level task (searching for human victims) with minimal input from human operators, by navigating in a complex, dynamic and environment, while avoiding potentially hazardous obstacles?

To solve these issues, an outdoor mobile robotic platform, as shown in figure 1, was equipped with a differential GPS system for accurate geo-registered positioning, and a stereo vision system. This stereo vision systems serves two purposes: 1) victim detection and 2) obstacle detection and avoidance. For semi-autonomous robot control and navigation, we rely on a behavior-based robot motion and path planner. In this paper, we present each of the three main aspects (victim detection, stereo-based obstacle detection and behavior-based navigation) of the general robot control architecture more in detail.

1. Victim Detection

Detecting victims lying on the ground using standard camera images is very different from standard person detection, which is a common research subject in the computer vision community. These standard person detection algorithms generally rely on face [7],[8] or upper body [5] detection, which provide powerful cues for reasoning bout a person's presence. However, these approaches assume that the person's face is clearly visible in the camera image and that the person is standing straight up, such that the upper body can be easily detected. Victims, however, do not tend to stand up. Moreover, in order to scan a large outdoor area rapidly, the field of view of the robot cameras is quite large, which means that a person's face only consists of a limited number of pixels. To achieve robust victim detection in these difficult outdoor conditions, the Viola-Jones algorithm [10] for Haar-features based template recognition was adapted to recognize persons lying on the ground.

The Viola-Jones method gives a visual object detection framework that is capable of rapidly processing images, while achieving high detection rates. There are three key aspects. The first is the introduction of an image representation called the *Integral Image*, which allows the features used by the detector to be computed quickly. The second is a learning algorithm, based on AdaBoost [2], which selects a small number of critical visual features and yields efficient classifiers. The third aspect is a method for

Figure 2. Detected Human Victims on Live Camera Images.

combining classifiers in a cascade, which allows background regions of the image to be quickly discarded while spending more computation on promising object-like regions. Viola and Jones originally applied this technique in the domain of face detection [10]. Their system yields face detection performance comparable to the best previous systems. For the victim-detection application, we adapted the Viola-Jones technique, by training the algorithm with bodies, lying on the ground.

To deal with the huge number of degrees of freedom of the human body and the camera viewpoint, the configuration space for human victims was reduced to victims lying face down and more or less horizontally in front of the camera. This case has been chosen because in real disasters this pose has the highest probability. The people try to protect their head and their ventral body parts which are the most vulnerable. Another reason is that in this position, the possible positions of the limbs form a relatively small pool comparing to the other cases. Also the orientation of the body must be considered because the legs have a different shape than the upper body and the head. To handle this, the sample images were taken with the both body orientations (left-to-right and right-to-left). To enlarge the data-set, the images were then later flipped horizontally and reused during the Haar-training.

Figure 2 illustrates the output of the victim detection module on test images from a live camera. As can be noted, the human victims lying on the ground are correctly identified, as visualized by the bounding rectangles drawn around the human bodies. Tests with real-time camera streams show that the correct detection rate of the algorithm is approximately 65%. This can be improved upon even further, using some post-processing technique (e.g. Kalman filtering) for integrating the results over multiple camera frames. Running the two victim detectors (one with left-to-right and one with right-to-left body orientation), the processing time is between 60 and 80 milliseconds, which means 13 to 16 frames per second.

2. Stereo-based Obstacle Detection

Detecting obstacles from from stereo vision images may seem simple, as the stereo vision system directly delivers rich depth information. However, from this depth image, it is not evident to distinguish the traversable from the non-traversable terrain, especially in outdoor conditions, where the terrain roughness and the robot mobility parameters must be taken into account. Our approach is based on the construction and subsequent processing of the *v-disparity* image [4], which provides a robust representation of the geometric content of road scenes. The v-disparity image is constructed by calculating a horizontal histogram of the disparity stereo image.

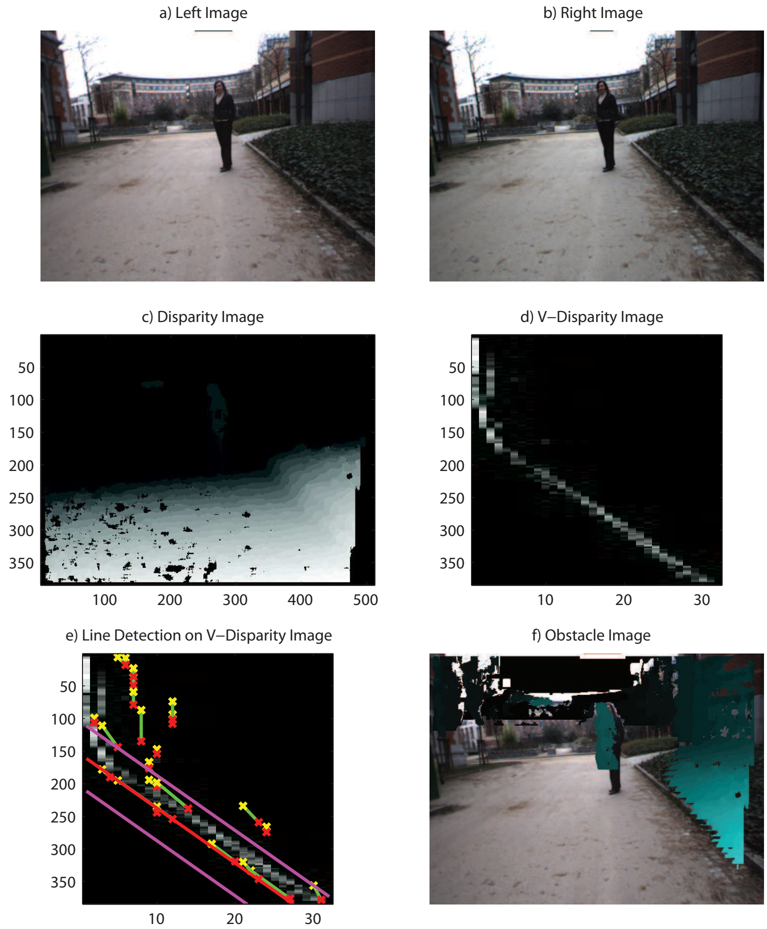

Figure 3. Different Steps of the Stereo-based Obstacle Detection Algorithm.

Consider 2 stereo frames, as shown in figure $3a$ and b, and a disparity image I_D, as shown in figure $3c$. Then, the v-disparity image I_V can be constructed by accumulating the points with the same disparity that occur on a horizontal line in the image. Figure $3d$ displays the v-disparity image I_V for the given input images.

The classification of the terrain in traversable and non-traversable areas goes out from the assumption that the majority of the image pixels are related to traversable terrain of the ground plane. The projection of this ground plane in the v-disparity image is a straight line, from the top left to the bottom right of the v-disparity image. Any deviations from this projection of the ground plane are likely obstacles or other non-traversable terrain items.

As such, the processing of the v-disparity image comes down to estimating the equation of the line segment in the v-disparity image, corresponding to the ground plane, as indicated by the red line in figure $3e$. Then, one must choose a single parameter which accounts for the maximum terrain roughness. As this parameter depends only on the robot characteristics, it only needs to be set once. This parameter sets the maximum offset in v-disparity space to be considered part of the ground plane. The two pink lines on figure $3e$ indicate the region in v-disparity space where pixels are considered part of a traversable region.

Any outliers are regarded as obstacles, which enables to compile an obstacle image I_O as displayed on figure $3f$. From figure $3f$, it is clear that non-traversable areas (the bushes) and obstacles (the person) are very well distinguished. It may be noted that the lower part of the legs of the person were not detected as obstacles. This is due to the choice of the threshold parameter for the ground plane, discussed above. After tests in multiple environments, we used a threshold parameter of 50, which offers a good compromise between a good detection rate and low false positive detection rate.

3. The Behavior-based Robot Navigation Architecture

Figure 4 illustrates the general robot control architecture, set up as a testbed for the algorithms discussed in this paper. The RobuDem robot used in this setup features 2 on-board processing stations, one for low-level motor control (Syndex Robot Controller), and another one for all the high-level functions. A remote robot control PC is used to control the robot and to visualize the robot measurements (color images, victim data) from a safe distance. All data transfer between modules occurs via TCP and UDP-based connections, relying on the CORBA [9] and CoRoBa [1] protocols. To increase the bandwidth and to assure the quality of service over the wireless link from the on-board high-level PC to the remote robot control PC, the use of the MailMan protocol over Wi-Max is investigated.

A behavior-based navigational architecture is used for semi-autonomous intelligent robot control. Behavior-based techniques have gained a widely popularity in the robotics community [3], due to the flexible and modular nature of behavior-based controllers, facilitating the design process. Following the behavior based formalism, a complex control task is subdivided into a number of more simple modules, called behaviors, which each describe one aspect of the sensing, reasoning and actuation robot control chain. Each behavior outputs an objective function, $o_1(\mathbf{x}), ..., o_n(\mathbf{x})$, which are multi-dimensional normalized functions of the output parameters, where $\mathbf{x} = (x_1, ..., x_n) \in R^n$ is an $n-$dimensional decision variable vector. The degree of attainment of a particular alternative $\mathbf{x}$, with respect to the k^{th} objective is given by $o_k(\mathbf{x})$.

Recall that the RobuDem robot is equipped with two main sensing abilities: a stereo vision system and a GPS system. The information from the stereo vision system is used threefold. First, the color images are sent over the wireless link, such that the human operator receives at all time a visual cue of the environment. Secondly, the (left) color image is sent to the victim detection module, discussed in section 1. The victim detection module will report any detected human victims back to the human operator at the remote control station. Third, the calculated stereo disparity image is sent to the obstacle detection module, discussed in section 2. From the obstacle map, a behavior is constructed to steer the robot away from obstacles.

The GPS system delivers accurate robot positioning information, which is sent to the operator at the remote control station. At the same time, this data is sent to a path planning module. From the robot control station, the human operator is able to compile a list of waypoints for the robot. The path planning module compares this list of waypoints with the robot position and calculates a trajectory to steer the robot to the first goal position in the list. The first point on this trajectory list is sent to a *GoToGoal* behavior module,

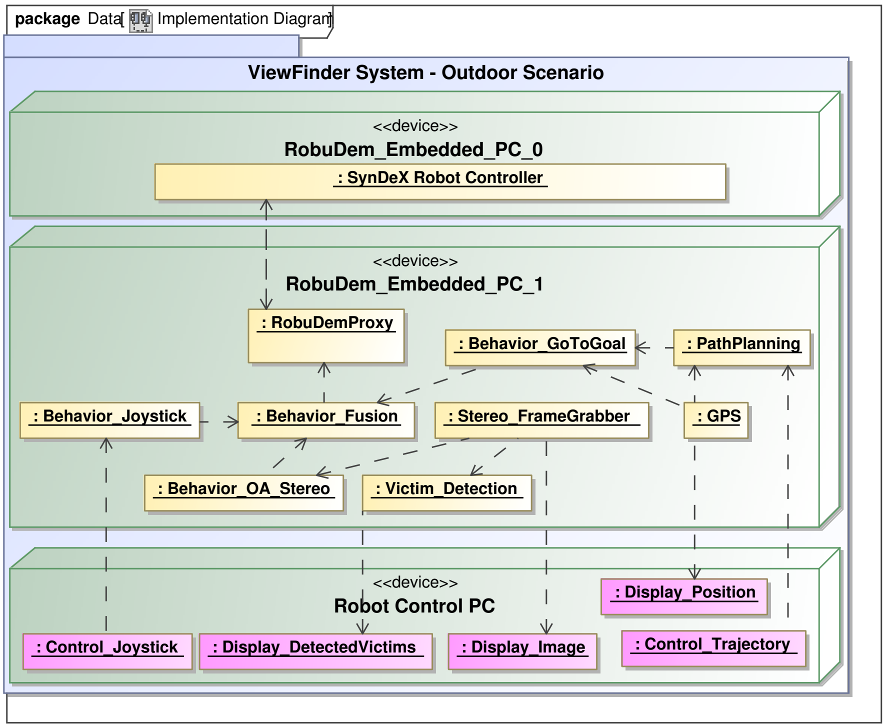

Figure 4. The Robot Control Architecture.

which aims to steer the robot to this point, as such executing the trajectory defined by the path planner.

In the case of robot control, the objective function of each behavior can be regarded as two-dimensional normalized function of robot steering velocity v and direction α. For this setup, three behaviors are defined which relate the abstract sensor information into robot actions. These three behaviors are:

1. *Obey Joystick Commands*. If desired, the human operator can control the robot by means of a joystick. The joystick commands are directly related to the robot steering angle and direction, so the transformation of the joystick control command into an objective function can be performed straightforward by calculating a two-dimensional Gaussian from the joystick input $(v_{Joystick}, \alpha_{Joystick})$:

$$o_{Joystick}(v, \alpha) = \frac{1}{\sqrt{(2\pi)^2 \sigma^4}} e^{-\left(\frac{\left(v - v_{Joystick}\right)^2}{2\sigma^2} + \frac{\left(\alpha - \alpha_{Joystick}\right)^2}{2\sigma^2}\right)} \tag{1}$$

2. *Obstacle Avoidance Using Stereo*. To drive the robot away from obstacles detected by the stereo vision system, the obstacle image I_O is analyzed. The depth values of pixels corresponding to obstacles are accumulated per vertical line in the image and the resulting function is inverted and normalized. This allows to deduce a function f of the viewing angle α as shown on figure 5. This func-

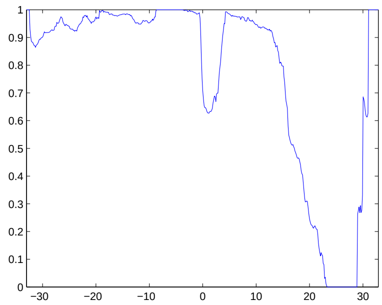

Figure 5. 1D Objective Function for Obstacle Avoidance from Stereo, corresponding to the input of figure 3.

tion can be regarded as a one-dimensional objective function for obstacle avoidance from stereo input, considering only the viewing / steering angle. It can be noted on figure 5, which corresponds to the input of figure 3, that for example the nearby bushes on the right side of the robot make a turn to the right less desirable. This one dimensional objective function can then be extended for velocity as well, using the following formulation:

$$o_{Stereo}\left(v,\alpha\right)=\frac{f\left(\alpha\right)}{1+\left|vf\left(\alpha\right)/c\right|} \tag{2}$$

3. *Go To Goals*. The goal seeking behavior is assigned two tasks. First, it points the robot to the goal position and it varies the velocity respective to the distance to the goal. This means the development of the objective function can be split up as $\mathbf{o}_{GoToGoal}(v,\alpha)=\mathbf{o}^{\alpha}_{GoToGoal}(\alpha).\mathbf{o}^{v}_{GoToGoal}(v)$. To calculate these objective functions, the (Euclidian) distance to the goal d_{goal} and heading to this goal θ are calculated from the current robot position given by the GPS system and the current waypoint given by the global path planner. The goal seeking behavior aims to minimize the difference between the robot heading α and the goal heading θ, which can be formulated as:

$$\mathbf{o}^{\alpha}_{GoToGoal}(\alpha)=\frac{1}{1+\left(\frac{\alpha-\theta}{\beta}\right)^{2}}\ . \tag{3}$$

with β the window size which is considered. $\mathbf{o}^{v}_{GoToGoal}(v)$ is set up such that the velocity is always high, with the exception that when the robot approches a goal position, the speed should be reduced. This is expressed as:

$$\mathbf{o}^{v}_{GoToGoal}(v)=\begin{cases}\left(\frac{v}{v_{\max}}\right)^{2} & \text{if} \quad d_{\text{goal}}>d_{\text{threshold}}\\ \frac{1}{1+\left(\frac{v}{v_{\max}}\right)^{2}} & \text{if} \quad d_{\text{goal}}<d_{\text{threshold}}\end{cases}\ . \tag{4}$$

These 3 behaviors must be fused together to form one consistent and globally optimal robot command, to be sent to the robot actuators. The performance of the behavior-based controller depends on the implementation of the individual behaviors as well as on the method chosen to solve the behavior fusion or action selection problem. We have chosen a method to solve the action selection problem, by formulating it as as a multiple objective decision making problem [6]. Mathematically, a multi-objective decision problem can be represented as finding the solution to $\arg\max_{\mathbf{x}} [o_1(\mathbf{x}), ..., o_n(\mathbf{x})]$. The method followed for solving the multiple objective decision making problem problem is the weighting method. This method is based on scalar vector optimization and is formulated in the following way:

$$\mathbf{x}^* = \arg\max_{\mathbf{x}\in X} \sum_{i=1}^{n} w_i o_i(\mathbf{x}). \tag{5}$$

where w_i are normalized weights such that $\sum_{i=1}^{n} w_i = 1$. The solution to equation 5, $\mathbf{x}^*(v^*, \alpha^*)$ defines the control command which is sent to the robot.

4. Conclusions

This paper has presented the development of a mobile outdoor robot with three main capabilities: automated victim detection, obstacle detection from stereo images and behavior-based semi-autonomous control. Using this control strategy, the robot is able to navigate on its own in an outdoor environment, navigating to a set of human-operator-defined way-points while avoiding obstacles. At the same time, the remote human operator is automatically alerted when human victims are detected.

In the future, research on this robotic platform will be intensified with the goal of equipping the robot with a chemical sensing unit and a visual simultaneous localization and mapping module. This will enable the robot to report the presence of toxic gasses to a remote operator, while mapping the incident site.

A robot with these capacities is a valuable assistant for crisis management teams, as it performs a potentially hazardous and potentially life-saving task: navigating through unknown terrain and mapping this terrain while finding human survivors and informing the human firefighters about the presence of toxic gasses in the environment.

References

[1] E. Colon, H. Sahli and Y. Baudoin, "CoRoBa, a multi mobile robot control and simulation framework", *International Journal of Advanced Robotic Systems*, Vol. 3, Number 1, pp. 073-078, March 2006.

[2] Y. Freund, R. E. Schapire: "A decision-theoretic generalization of on-line learning and an application to boosting", *Journal of Computer and System Sciences*, Volume 55 (1) pp. 119.139, August 1997.

[3] J.L. Jones, "Robot Programming: A practical guide to Behavior-Based Robotics", 2004.

[4] R. Labayrade, D. Aubert and J.P. Tarel, "Real Time Obstacle Detection on Non Flat Road Geometry through 'V-Disparity' Representation", *Proceedings of IEEE Intelligent Vehicle Symposium*, June 2002.

[5] A. S. Micilotta, E.J. Ong and Richard Bowden, "Real-Time Upper Body Detection and 3D Pose Estimation in Monoscopic Images", *Computer Vision ECCV 2006*, pp. 139-150, 2006.

[6] P. Pirjanian. "Behavior coordination mechanisms - state-of-the-art". Tech. Rep. IRIS-99-375, Institute of Robotics and Intelligent Systems, School of Engineering, Univ. of South California, Oct. 1999.

[7] H. Rowley, S. Baluja and T. Kanade, "Neural Network-Based Face Detection", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 20, No. 1, pp. 23-38, January, 1998.

[8] H. Schneiderman and T. Kanade, "A Statistical Model for 3D Object Detection Applied to Faces and Cars", *IEEE Conference on Computer Vision and Pattern Recognition*, June, 2000.

[9] D. Slama, J. Garbis and P. Russell, "Enterprise CORBA", Prentice Hall, 1999

[10] P. Viola and M. Jones: "Robust Real-time Object Detection", *2nd International Workshop on Statistical and Computational Theories of Vision-Modeling, Learning, Computing and Sampling*, Canada, 2001.