Using visual perception for controlling an outdoor robot in a crisis management scenario

Conference Paper · September 2010

CITATIONS
0

READS
1,226

4 authors:

Geert De Cubber
Royal Military Academy
89 PUBLICATIONS **406** CITATIONS
SEE PROFILE

Daniela Doroftei
Royal Military Academy
46 PUBLICATIONS **230** CITATIONS
SEE PROFILE

Sid Ahmed Berrabah
Royal Military Academy
33 PUBLICATIONS **127** CITATIONS
SEE PROFILE

Yvan Baudoin
Royal Military Academy
55 PUBLICATIONS **229** CITATIONS
SEE PROFILE

Using visual perception for controlling an outdoor robot in a crisis management scenario

G. De Cubber, D. Doroftei, S.A. Berrabah, Y. Baudoin

Royal Military School of Belgium, Department of Mechanics,

30 Avenue De La Renaissance, B1000 Brussels, Belgium

geert.de.cubber@rma.ac.be

Abstract

Crisis management teams (e.g. fire and rescue services, anti-terrorist units ...) are often confronted with dramatic situations where critical decisions have to be made within hard time constraints. Therefore, they need correct information about what is happening on the crisis site. In this context, the View-Finder projects aims at developing robots which can assist the human crisis managers, by gathering data. This paper gives an overview of the development of such an outdoor robot. The presented robotic system is able to detect human victims at the incident site, by using vision-based human body shape detection. To increase the perceptual awareness of the human crisis managers, the robotic system is capable of reconstructing a 3D model of the environment, based on vision data. Also for navigation, the robot depends mostly on visual perception, as it combines a model-based navigation approach using geo-referenced positioning with stereo-based terrain traversability analysis for obstacle avoidance. The robot control scheme is embedded in a behavior-based robot control architecture, which integrates all the robot capabilities. This paper discusses all the above mentioned technologies.

Keywords

Visually guided robots, Dense Structure from Motion, Behavior-based Robot Control, Intelligent Mobile Outdoor Robots, Crisis Management

1. Introduction

1.1. Goal and problem statement

When confronted with a large crisis, the human crisis managers require a complete overview of the crisis site is necessary to take correct decisions. 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 crashed airplane ...) 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. The design requirements for such a robotic crisis management system give rise to four main problems which need to be solved for the successful development and deployment of such a mobile robot:

1. How can the robot automatically detect human victims, even in difficult outdoor illumination conditions?
2. How can the robot, which needs to navigate autonomously in a totally unstructured and unknown environment, auto-determine the suitability of the surrounding terrain for traversal?
3. How can the robotic system increase the perceptual awareness of the human crisis managers?
4. How can the robot be made semi-autonomous, such that the human crisis managers are not overloaded with the task of controlling the robot?

In response to the first question, we present an approach to achieve robust victim detection in difficult outdoor conditions, by going out from the Viola-Jones algorithm for Haar-features based template recognition and adapting it to recognize victims. Victims are assumed to be human body shapes lying on the ground. The algorithm tries to classify visual camera input images into human body shapes and background items. This approach is further explained in section 2.1.

The second problem which is stated above is that of the traversability estimation. This is a challenging problem, as the traversability is a complex function of both the terrain characteristics, such as slopes, vegetation, rocks, etc. and the robot mobility characteristics, i.e. locomotion method, wheel properties, etc. In section 2.2, we present an approach where a classification of the terrain in the classes "traversable" and "obstacle" is performed using only stereo vision as input data.

In response to the third question, we propose an image-based 3D reconstruction technique, enabling to reconstruct a global 3D model of the environment, as seen by the robot. Using this global 3D model, valuable information (presence of victims, dangerous gasses …) can be visualized to the crisis managers in a user friendly interface. The proposed 3D reconstruction methodology is further explained in section 3.

The final question raises the important issue that any robotic system should not increase the cognitive load for the human crisis managers. These human crisis managers prefer a scenario where they designate a working area to the robot, or select a set of interesting locations to investigate, and then the robot should execute this high-level task. Therefore, it is required that the robot can navigate to geo-referenced locations on a map. A behavior based control paradigm was chosen as a control mechanism to combine all robot capabilities in a comprehensive and modular framework, such that the robot 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, using stereo vision as a main source of sensor information. The behavior based control architecture is further detailed in section 3 of this paper.

1.2. System description

The "ROBUDEM" outdoor mobile robotic platform which was developed in the framework of the View-Finder project is shown on Figure 1. It is equipped with a GPS system for accurate geo-registered positioning and a stereo vision system. Next to these primary sensors it also disposes of 5 sonar sensors for collision avoidance and an Inertial Measurement Unit (IMU) for orientation measurement. For processing, the robot is equipped with two PC's: one featuring a Real-Time Linux operating system for low-level motor control, and one 1.6GHz Dual Core Windows PC, dealing with all high-level processing tasks. It is important to note that the developed robot must be seen as a research platform, designed to evaluate the presented technologies, not as an actual prototype of a real crisis management robot. For this purpose, the current platform is too large, too heavy and not fire and weatherproof enough. However, it presents an excellent platform to test the technologies presented in this paper and which can then be later integrated on a more robust real crisis management robot.

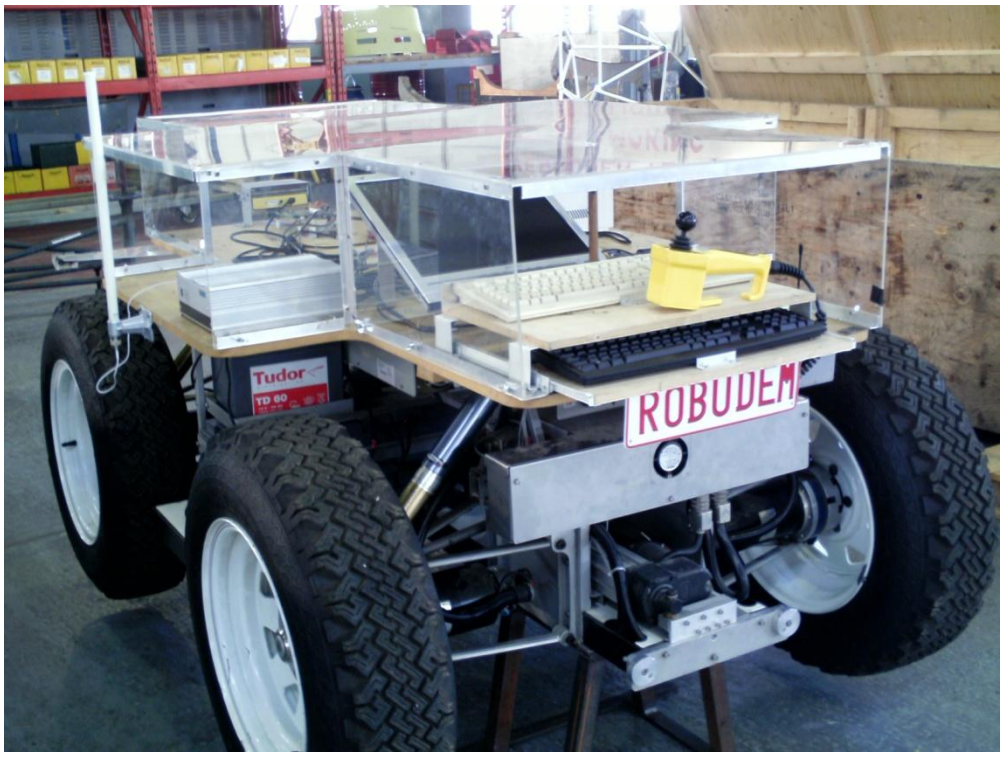

Figure 1: The ROBUDEM robot used as a test platform for the presented technologies.

2. Visual perception

The Robudem robot used for the View-Finder project relies on vision as its primary sensing modality. Therefore, the amount of information which can be extracted from the measurements acquired by the on-board stereo camera system must be maximized. To this extent, multiple processing cues for the visual data are established. An important aspect of all these processing cues is that the information they deliver must be available in real-time, or near real-time. This constraint limits the complexity of the applied algorithms and calls for a balanced compromise between the quality of the output and the required processing time.

2.1. Human victim detection

Automated human victim detection is a very difficult task, especially in complex, unstructured environments. In order to detect a human victim, one or more physical parameters of the victim need to be perceived by a sensor. These physical parameters can be (Burion, 2004): voice, temperature, scent, motion, skin color, face or body shape. Here, we present an approach to achieve robust victim detection in difficult outdoor conditions.

The basis for this work is a learning-based object detection method, proposed by Viola and Jones (Viola, 2004). Viola and Jones originally applied this technique in the domain of face detection. Their system yields face detection performance comparable to the best previous systems and is considered the fastest and most accurate pattern recognition method for faces in monocular grey-level images.

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 re-used during the Haar-training.

For database training, 800 positive scenes were recorded, featuring human victims in several orientations and under varying illumination. These images were taken with an on-board stereo camera system. Furthermore 500 pairs of negative images were recorded outside and 100 pairs inside. These images contain no humans but the variety of the background is high, such that the learning method sets up good thresholds.

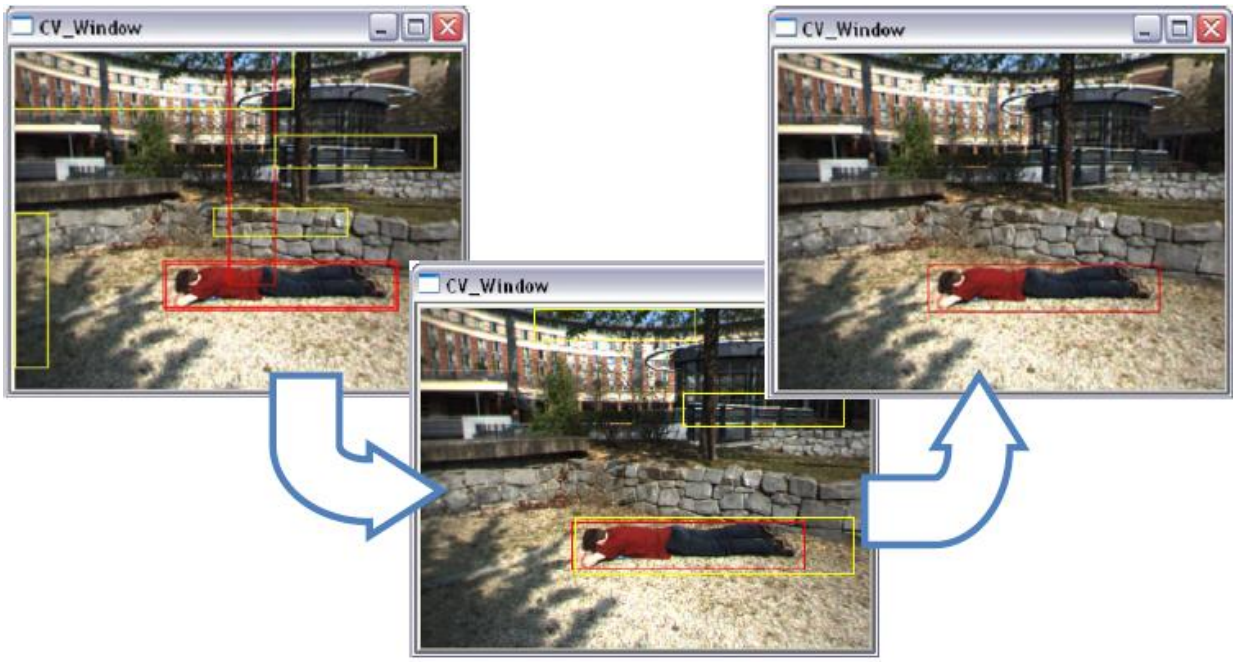

Figure 2: An example test image for Victim Detection

Theoretically, the classifier for victim detection has a 100% detection rate and less than 10-6 % false alarm rate. Of course, this is only true in the case of the positive and negative sample images. With new test images - which were taken in similar illumination conditions as the sample images but in different positions - the correct detection rate was approximately 65%.
Figure 2 shows the result of the victim detection algorithm. The red rectangles are the hits of the detector for the victims whose head is at the left, the yellow ones for those whose head is at the right. In the first image of Figure 2, the victim was correctly found besides of a lot of false positives. These false alarms are eliminated by merging the adjacent rectangles of correct posture. The processing time for running the victim detector is between 60 and 80 milliseconds, which means that we are able to achieve 13 to 16 frames per second. This is a very good result, as it allows near real-time reactions in the robot control scheme and it also allows integrating the results of multiple detection runs over time by means of a tracking scheme, to enhance the detection rate and reduce the false positive rate.

2.2. Terrain-traversability estimation

Terrain traversability analysis is a research topic which has been in the focus of the mobile robotics community in the past decade, inspired by the development of autonomous planetary rovers and, more recently, the DARPA Grand Challenge.
In this paper, we present an approach where a classification of the terrain in the classes "traversable" and "obstacle" is performed using only stereo vision as input data. In a first step, high-quality stereo disparity maps are calculated by a fast and robust algorithm (Scharstein, 2002). Using this stereo depth information, the terrain classification is performed. Detecting obstacles from stereo vision images may seem simple, as the stereo vision system can provide rich depth information. However, from the 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.

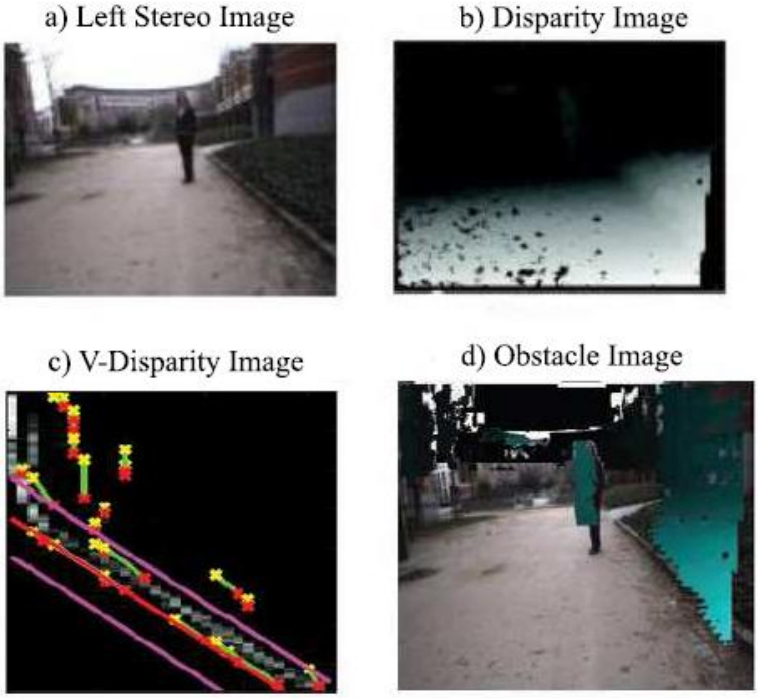

Figure 3: Terrain Traversability Estimation: a) Left stereo image; b) Disparity image; c) V-Disparity image after Hough transform; d) Obstacle image.

Our approach is based on the construction and subsequent processing of the v-disparity image (Labayrade, 2002), 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. Consider 2 stereo frames, as shown in Figure 3a, and the computed disparity image I_D, as shown in Figure 3b. 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 3c 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. This is done by performing a Hough transform on the v-disparity image and searching for the longest line segment. The red line in Figure 3c indicates the largest line segment, corresponding to the ground plane.
Finally, 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 in Figure 3c indicate the region in v-disparity space where pixels are considered part of a traversable region. Terrain corresponding to pixels in v-disparity space in between the two pink lines is considered traversable, while any outliers are regarded as obstacles, which enables to compile an obstacle image I_O. The result of this operation can be judged from Figure 3d, showing the obstacle image. This is a version of the color input image, where false color data corresponding to the disparity is superposed for pixels classified as belonging to non-traversable terrain.
It may be noted that the lower part of the legs of the person standing in front of the robot were not detected as obstacles. This is due to the choice of the threshold parameter for the ground plane. 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. Visual 3D reconstruction

When confronted with a large crisis, the crisis management teams require a global overview of the crisis scene. In practical situations, however, it is near impossible to obtain such a global overview, due to the abundance of information coming from different sources and the lack of a global model of the crisis scene where all this information can be nicely visualized upon. In this section, we propose an automated 3D reconstruction approach for building a global 3D model. This 3D reconstruction approach is based on dense structure from motion (SFM) recovery from images captured by a camera on-board a semi-autonomous crisis management robot. Dense structure from motion algorithms aim at estimating a 3D location for all camera image pixels.

There are multiple approaches towards dense structure from motion. The most modern dense structure from motion algorithms minimize the optical flow constraint and enforce smoothness in the depth field in a variational framework. However, due to the noisiness of the optical flow and due to projection ambiguities, these algorithms are still not very robust when confronted with unconstrained 3D camera motion and changing illumination conditions. One could argue that these problems are due to the fact that dense structure from motion is a relatively new field of research that emerged recently due to the rise in computing power.

Sparse structure from motion, on the other hand, is a more mature research domain, which dates back to the early work of Longuet-Higgins (Longuet-Higgins, 1981). Through the years, sparse structure from motion algorithms have been optimized and made more robust, notably by the work of Hartley and Zissermann (Hartley, 2004).

To address the classical dense structure from motion shortcomings, we adopt a dual approach for dense structure estimation (De Cubber, 2010). The approach combines the strength of the more robust feature-based structure from motion approaches with the spatial coherence of dense reconstruction algorithms. Dense reconstruction is regarded as a high-dimensional data fusion problem with as inputs the camera motion parameters and 3D coordinates of feature points estimated by sparse reconstruction and dense optical flow. The base constraint of the variational approach is the traditional image brightness constraint, but parameterized for the depth using the 2-view geometry. This estimation of the geometry, as expressed by the fundamental matrix, is automatically updated at each iteration of the solver. A regularization term is added to ensure good reconstruction results in image regions where the data term lacks information. An automatically updated regularization term ensures an optimal balance between the data term and the regularization term at each iteration step. A semi-implicit numerical scheme was set up to solve the dense reconstruction problem. The solver goes out from am initialization process which fuses optical flow data and sparse feature point matches.

The developed methodology is capable of estimating a high quality 3D reconstruction of a natural scene (as presented in section 5), which makes it a valuable tool for human crisis management teams. Indeed, the results obtained during a real crisis management exercise, show that the visual models outputted by the presented method, can increase the situational awareness of the human crisis managers by integrating localized information on the 3D model.

4. Behavior-based robot control

Figure 4 illustrates the general robot control architecture, set up as a test bed 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 from a safe distance. All data transfer between modules occurs via TCP and UDP-based connections, relying on the CoRoBa (Colon, 2006) protocol.
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 (Jones , 2004), 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, $o(x)$, which is a multi-dimensional normalized function of the output parameters, where x is an n-dimensional decision variable vector. The degree of attainment of a particular alternative x, with respect to the k^{th} objective is given by $o_k(x)$.
A first behavior implemented on the robot is a goal seeking behavior. 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 goal positions in the list. The first point on this trajectory list is sent to a *GoToGoal* behavior module, which aims to steer the robot to this point, as such executing the trajectory defined by the path planner.

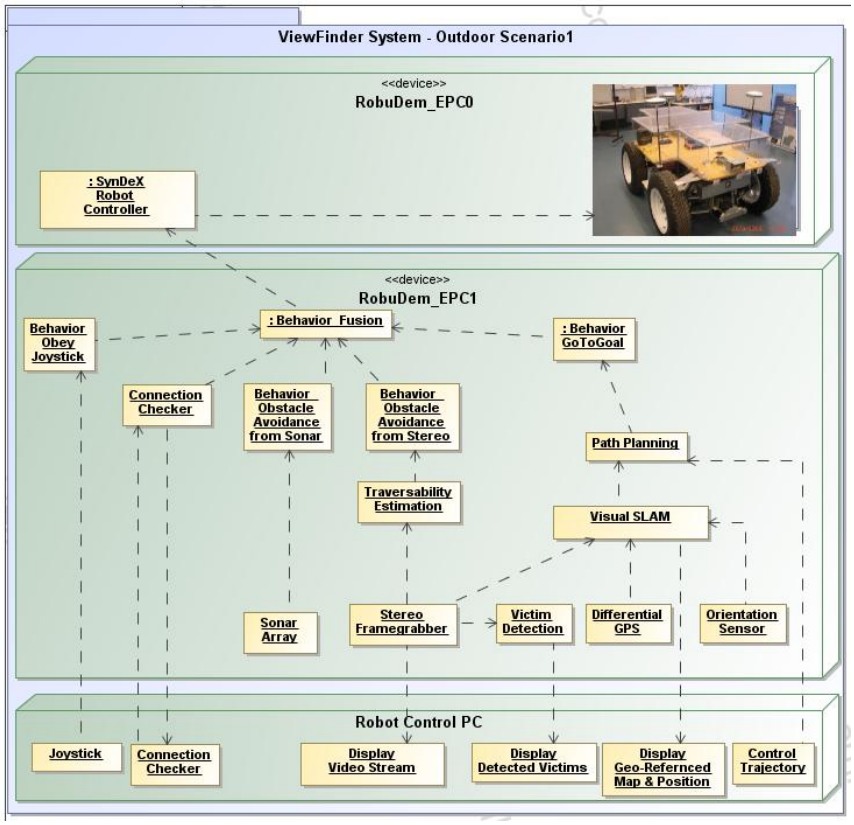

Figure 4: Robot Control Architecture

The ROBUDEM robot uses vision as a primary sensing modality and is therefore equipped with a stereo vision system. The information from this stereo vision system is used threefold:

1. The (left) color image is sent to a victim detection module, as presented in section 2.1. The victim detection module will report any detected human victims back to the human operator at the remote control station. For now, it is up to the user to decide how to react to the victim detection event, there is no automated response of the robot.
2. The color images are sent over the wireless link, such that the human operator receives at all time a visual cue of the environment. This is absolutely necessary when the robot is operating under tele-operation mode. In tele-operation mode, a human operator can control the robot by means of a joystick. Therefore, a first behavior is set up such that the robot obeys to the joystick commands.
3. The stereo data is processed by a terrain traversability estimation module, as presented in section 2.2. The obstacle image, resulting from this process, is analyzed and an *ObstacleAvoidance* behavior is set up to drive the robot away from any obstacles.

The specific design of each of these behaviors is out of the scope of this paper and is also not of primordial importance, as it uses classical methods of designing objective functions for behaviors. As usual in behavior-based control, the more important question is how to combine the different behaviors to come to a globally optimal control command for the robot to execute. For this, we have used the approach proposed by Doroftei et al. in (Doroftei, 2009). Their methodology consists of an extension of the classical goal-programming method, by integrating it with an approach based on reliability analysis. This methodology has the advantage that it incorporates direct information from the system under control into the control process, while taking into account a decision maker's preferences. As such, a globally optimal control command can be estimated, which is used to steer the different robot actuators.

5. Integrated results of a live crisis management exercise

The presented technologies were tested during an integrated crisis management exercise, where an airplane crash was simulated. During this exercise, a semi-autonomous robot was asked by the firefighters to search for human survivors (in this case: the pilot who ejected from the airplane before the crash) near the incident site and while doing this, it was requested to build a 3D model of the environment. Figure 5 shows some images taken by the robot on-board camera during this validation test, whereas Figure 6 shows the reconstructed 3D model.

Figure 5: Some frames shot by the semi-autonomous robot during the crisis management exercise and an external view

The 3D model of Figure 6 shows a good resemblance to the physical nature of the environment and all required features can be identified: the ground plane, the bunker in the back, the canopy... As also the motion of the camera (which is fixed on the robot) is reconstructed using the presented methodology, the robot can be positioned in the virtual environment. As an example of how this 3D model can be efficiently used by crisis management teams, the 3D model of Figure 6 also indicates the position of a human survivor. The presence of the human survivor was detected by the human victim detection algorithm, as presented in section 2.1 and this information was fused with the 3D information obtained through the presented depth reconstruction algorithm to locate and visualize the victim in the 3D model. The visualization of the virtual 3D scene with added localized information, as presented by Figure 6, provides a powerful tool for the human crisis management teams to augment their situational awareness without increasing the cognitive load too much, as the whole process of data acquisition by the robot and processing by the presented algorithm is automated.

Figure 6: Reconstructed 3D model of the environment, showing the camera/robot position and an indication of the presence of human survivors

6. Conclusions and future work

The integration test discussed above can be called successful, as the was able to scan the designated zone without any major problems. However, to come to a robotic system which can be practically deployed, there are still a number of issues to be solved:

- The response time must be drastically reduced, by better integration of robot services;
- The false detection rate of the victim detector must be reduced by embedding the detector in a tracking scheme;
- The terrain traversability estimation must be tested with different types of terrain;
- The 3D scene reconstruction algorithm must be made faster, such that the human operators can be presented a 3D view of the environment in near real-time;
- An active vision system with a larger field of view should be used;
- A practical robot prototype must be lighter, smaller, and more fire and weatherproof.

In the near future, the robot will also be equipped with a gas-sensor. This is of particular interest for fire fighters, as the presence of toxic or highly explosive chemicals at the incident site is a very important factor to assess when deciding to send in a team of human fire fighters.

Despite these obvious shortcomings which leave the way for future work, the current ROBUDEM robotic system very well succeeds in fulfilling the design requirements set up at the beginning of this project: 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. If required, a remote human operator can still take control of the robot via the joystick, but in normal operation mode, the robot navigates autonomously to a list of waypoints, while avoiding obstacles (thanks to the stereo-based terrain traversability estimation) and while searching for human victims.

References

Burion, S. (2004) Human detection for robotic urban search and rescue. Master's thesis.

Colon, E., Sahli, H., Baudoin, Y. (2006) Coroba, a multi mobile robot control and simulation framework. Intl.Journal of Advanced Robotic Systems, 3(1), 73-78.

De Cubber, G. (2010) Variational Methods for Dense Depth Reconstruction from Monocular and Binocular Video Sequences, PhD. Thesis

Doroftei, D., De Cubber, G., Colon, E., Baudoin, Y. (2009) Behavior based control for an outdoor crisis management robot. IARP Workshop on Robotics for risky interventions and Environmental Surveillance-Maintenance, Brussels, Belgium.

Hartley, R., Zisserman, A. (2004) Multiple View Geometry in Computer Vision. Cambridge University Press, New York, NY, USA, Second edition

Jones, J. (2004) Robot programming: A practical guide to behavior-based robotics

Labayrade, R. Aubert, D., Tarel, J.P. (2002) Real time obstacle detection on non-flat road geometry through v-disparity representation. Intelligent Vehicles Symposium, Versailles.

Longuet-Higgins, H.C. (1981) A computer algorithm for reconstructing a scene from two projections. Nature, 293(5828):133–135, 1981.

Scharstein, D., Szeliski, R. (2002) A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. Intl. Journal of Computer Vision, 47(1):7-42.

Viola, P., Jones, M. (2004) Robust real-time face detection. Intl. J. of Computer Vision, 57(2):137-154.