USARSim: Providing a Framework for Multi-robot Performance Evaluation

S. Balakirsky, C. Scrapper
NIST
Gaithersburg, MD, USA
stephen.balakirsky@nist.gov,
chris.scrapper@nist.gov

S. Carpin
International University Bremen
Bremen, Germany
s.carpin@iu-bremen.de

M. Lewis
University of Pittsburgh
Pittsburgh, PA, USA
ml@sis.pitt.edu

***Abstract*:** Research efforts in urban search and rescue robotics have grown substantially in recent years. Two important robotic competitions (a robot physical league and a high-level infrastructure simulation league) were established in 2001 under the RoboCup umbrella to foster collaboration amongst institutions and to provide benchmark test environments for system evaluation. While these leagues play valuable roles, a significant gap existed between simulating disaster infrastructure and implementing agent behaviors on real hardware. In this paper we describe a software simulation framework intended to be a bridge between these communities. The framework allows for the realistic modeling of robots, sensors, and actuators, as well as complex, unstructured, dynamic environments. Multiple heterogeneous agents can be concurrently placed in the simulation environment thus allowing for team or group evaluations. This paper presents a description of the simulation along with results from the RoboCup 2006 Virtual Robot Competition in which it was used and a roadmap of the framework's future directions.

Keywords: performance evaluation, simulation, USARSim, robotics

I. Introduction

Research in robotics for Urban Search and Rescue (USAR) has recently experienced vigorous development. USAR offers a unique combination of engineering and scientific challenges in a socially relevant application domain [4]. The broad spectrum of relevant topics attracts the attention of a wide group of researchers, with expertise as diverse as advanced locomotion systems, sensor fusion, cooperative multi-agent planning, human-robot interfaces and more.

The contest schema adopted by the RoboCup Rescue community, with the distinction between the real robots competition and the simulation competition, captures the two extremes of this growing community. The real robots competition is pushing the state of the art in robot mobility by challenging teams to perform in a room sized environment. These operations include tasks such as:

- autonomously negotiating compromised and collapsed structures,
- finding victims and ascertaining their condition,
- producing practical maps of victim locations,
- delivering sustenance and communications to victims,
- identifying hazards, and
- providing structural shoring.

The simulation competition's main purpose, by contrast, is to provide emergency decision support by integrating disaster information, prediction, planning, and human interfaces. The Version 0 simulator included simulations of building collapses, road blockages, spreading fire, and traffic. The competing teams must deploy scarce resources to address a dynamic disaster spreading over multiple city blocks. Both competition settings allow teams to be objectively evaluated in a challenging and realistic environment while providing a test arena for the development of performance metrics and standards for mobile robots.

Looking back at past RoboCup events, tremendous progresses in a short period of time has characterized both communities. In 2002, the real rescue robots competition was described as a competition where the complexity of the problem caused most researchers to use tele-operated robots [1]. In the simulation competition, emphasis was placed on the inter-agent communication models adopted [7]. The huge gap between these two extremes is evident.

Only two years later [5], the real robot competition saw the advent of teams with three dimensional mapping software, intelligent perception, and the first team with a fully autonomous multi-robot system. Within the simulation competition, teams exhibited cooperative behaviors, special agent programming languages, and learning components. With these strong gains, it is evident that relevant techniques will soon begin to migrate between the competitions. Nevertheless, certain logistic obstacles still prevent a seamless and profitable percolation of ideas and knowledge.

At RoboCup 2005, USARSim was selected as the software infrastructure for a new competition that fits between the physical and agent competitions. During the 2006 competition, eight teams from four continents competed in an indoor/outdoor city block sized virtual arena. The rest of this paper broken down as follows: Section 2 describes a short overview of the USARSim framework; Section 3 outlines the competition and performance metrics under which the teams were judged; finally, Section 4 presents lessons learned and tentative overview of rule and procedure changes for next year's competition.

II. USARSIM FRAMEWORK

The current version of USARSim is based on the UnrealEngine2 game engine that was released by Epic Games as part of Unreal Tournament 2004[1]. This engine may be inexpensively obtained by purchasing the Unreal Tournament 2004 game. The engine handles most of the basic mechanics of simulation and includes modules for handling input, output (3D rendering, 2D drawing, and sound), networking, physics and dynamics. Multiplayer games use a client-server architecture in which the server maintains the reference state of the simulation while multiple clients perform the complex graphics computations needed to display their individual views. USARSim uses this feature to provide controllable camera views and the ability to control multiple robots. In addition to the simulation, a sophisticated graphical development environment and a variety of specialized tools are provided with the purchase of Unreal Tournament.

The USARSim framework builds on this game engine and consists of

- standards that dictate how agent/game engine interaction is to occur,
- modifications to the game engine that permit this interaction
- an Application Programmer's Interface (API) that defines how to utilize these modifications to control an embodied agent in the environment
- 3-D immersive test environments.

In order to provide a standardized external interface, all units of measurement used in the USARSim API meet the International System of Units (SI) standard conventions. SI Units are a National Institute of Standards and Technology (NIST) developed convention that is built on the modern metric system, and is recognized internationally. For coordinate systems, USARSim leverages the previous efforts of the Society of Automotive Engineers, who published a set of standards for vehicle dynamics called *Vehicle Dynamic Terminology* [6]. This set of standards is recognized as the American National Standard for vehicle dynamics and contains a comprehensive set of standards that describes vehicle dynamics through illustrated pictures of coordinate systems, definitions, and formal mathematical representations of the dynamics. Finally, the messaging protocol, including the primitives, syntax, and the semantics are defined as part of the API.

When an agent is instantiated through USARSim, three basic classes of objects are created that provide for the complete control of the agent. These include robots, sensors, and mission packages and are defined as part of the API to USARSim. For each class of objects there are class-conditional messages that enable a user to query the component's geography and configuration, send commands, and receive status and data. Permissible calls into the game engine and complete details on the API may be found in the USARSim Reference Manual[2].

It is envisioned that researchers will utilize this framework to perfect algorithms in the areas of:

- Autonomous multi-robot control
- Human, multi-robot interfaces
- True 3D mapping and exploration of environments by multi-robot teams
- Development of novel mobility modes for obstacle traversal
- Practice and development for real robots that will compete in the physical league

III. VIRTUAL ROBOT COMPETITION

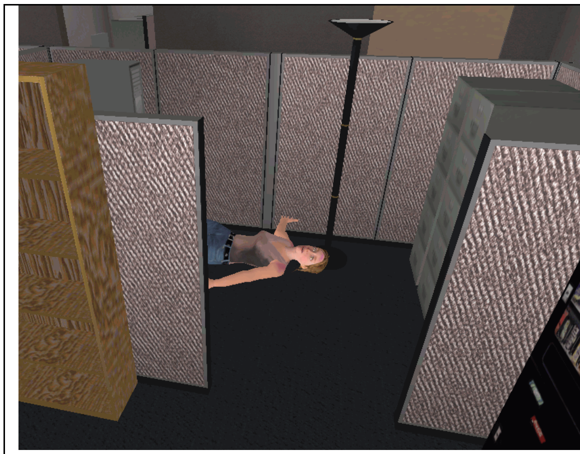
The RoboCup Rescue Virtual Competition is the third competition running under the RoboCupRescue Simulation League umbrella. It utilizes the USARSim framework to provide a development, testing, and competition environment that is based on a realistic depiction of a disaster scenario. It has been previously stated [2,3] that the Virtual Robots competition should serve the following goals:

- Provide a meeting point between the different research communities involved in the RoboCupRescue Simulation league and the RoboCupRescue Robot league. The two communities are attacking the same problem from opposite ends of the scale spectrum (city blocks vs. a small rubble area) and are currently far apart in techniques and concerns. The Virtual Competition offers close connections to the Robot league, as well as challenging scenarios for multi-agent research. The scenarios for the 2006 competition were chosen to highlight these connections. They were based on an outdoor accident scene, and an indoor fire/explosion at an office building. These scenarios included real-world challenges such as curbs, uneven terrain, multi-level terrain (i.e. the void space under a car), maze-like areas, stairs, tight spaces, and smoke. An exact copy of one of the RobCupRescue Robot league arenas was also included in the office space, and elements of other arenas were scattered

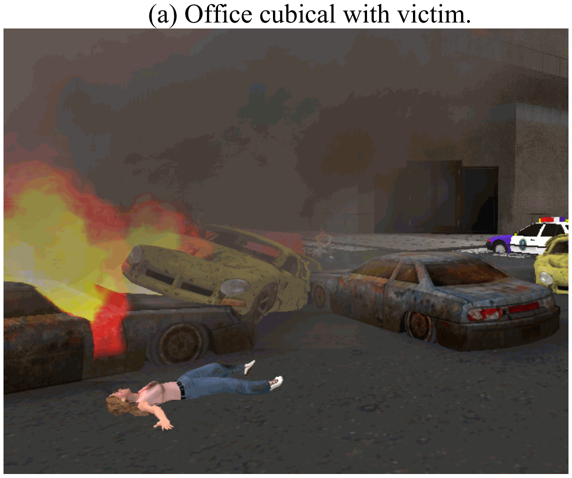
[1] Certain commercial software and tools are identified in this paper in order to explain our research. Such identification does not imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the software tools identified are necessarily the best available for the purpose.

[2] The reference manual may be found in the file releases area of the USARSim sourceforge site (http://sourceforge.net/project/showfiles.php?group_id=145394&package_id=180746&release_id=407397)

throughout the environment. The area was far too large to be explored by a single agent in the time permitted (20 minutes) and thus the use of multi-agent teams was beneficial. Accommodations were provided in the worlds to assist less capable (in terms of mobility) robotic systems. For example, wheel-chair ramps were provided that allowed for alternative access around stairs. Snap shots of small sections of these environments may be seen in Figure 1.



(a) Office cubical with victim.



(b) Outdoor traffic accident.

Figure 1: Representative snapshot of a USARSim indoor (a) and outdoor (b) sceen.

- Lower enter barriers for newcomers. The development of a complete system performing search and rescue tasks can be overwhelming. The possibility to test and develop control systems using platforms and modules developed by others makes the startup phase easier. With this goal in mind, the open source strategy already embraced in the other competitions is fully supported in the RobocupRescue Simulation league. Software from this year’s top teams has already been posted on the web.
- Let people concentrate on what they can do better. Strictly connected to the former point, the free sharing of virtual robots, sensors, and control software allows people to focus on certain aspects of the problem (victim detection, cooperation, mapping, etc), without the need to acquire expensive resources or develop complete systems from scratch. In order to help people determine if they really can “do better”, performance metrics were applied to the competing systems.

IV. Performance Metrics

For the 2006 competition, it was decided that performance would be measured in terms of information provided about victims, amount of area explored, and map quality. The primary goal of the competition was to locate victims in the environment. However, what does it meant to “locate” a victim? Several interpretations exist ranging from simply requiring a robot to be in proximity of a victim (e.g. drive by the victim) to requiring the robot to employ sensor processing to recognize that a victim is located near-by (e.g. recognize a human form in a camera image). It was decided that robots should be required to be “aware” of the presence of a victim, but that requiring every team to have expertise in image processing was against the philosophy of lowering entry barriers. Therefore, a new type of sensor; a victim sensor, was introduced.

This sensor was based on Radio Frequency Identification Tag (RFID) technology. False alarm tags were scattered strategically in the environment, and each victim contained an embedded tag. At long range (10 m), a signal from the tag was readable when the tag was in the field of view (FOV) of the sensor. At closer range (6 m), the sensor would report that a victim or false alarm was present. At even closer range (5 m) the ID of the victim would be reported. Finally, at the closest range (2 m), the status of the victim was available. Points were subtracted for reporting false alarms, and were awarded for information on the victims. Bonus points were awarded for including an image of the victim with the report.

As the robots were exploring the environment, their poses (on a 1 s interval) and any collisions between the robots and victims were automatically logged. The pose information was fed into a program that automatically computed the amount of area that was covered by the robotic teams. This figure was normalized against the expected explored area for the particular run, and points were awarded accordingly. The collision information was used as an indication of suboptimal navigation strategies that should be penalized. Another parameter that was used to determine the overall score was the number of human operators that were needed to control the robots. The idea was borrowed from the Physical Robots competition with the intent of promoting the deployment of fully autonomous robot teams, or the development of sophisticated human-robot interfaces that allow a single operator to control many agents.

The final area that was judged during the competition was map quality. The map quality score was based on several components.

- Metric quality – The metric quality of a map was scored automatically by examining the reported locations of "scoring tags". Scoring tags are RFID tags that report their relative location to a robot and then disappear. A requirement of the competition was for the teams to report the global coordinates of these tags at the conclusion of each run. The automatic scoring program then analyzed the deviation of the perceived locations from the actual locations.
- Multi-vehicle fusion – Teams were only permitted to turn in a single map file. Those teams that included the output from multiple robots in that single map were awarded bonus points.
- Attribution – One of the reasons to generate a map is to convey information. This information is often represented as attributes on the map. Points were awarded for including information on the location, name, and status of victims, the location of obstacles, the paths that the individual robots took, and the location of RFID scoring tags.
- Grouping – A higher order mapping task is to recognize that discrete elements of a map constitute larger features. For example the fact that a set of walls makes up a room, or a particular set of obstacles is really a car. Bonus points were awarded for annotating such groups on the map.
- Accuracy – An inaccurate map may make a first responder's job harder instead of easier. Points were assessed based on how accurately features and attributes were displayed on the map.
- Skeleton quality – A map may be inaccurate in terms of metric measurements (a hallway may be shown to be 20 m long instead of 15 m long), but may still present an accurate skeleton (there are three doors before the room with the victim). The category allowed the judges to award points based on how accurately a map skeleton was represented.
- Utility – One of the main objectives of providing a map was to create the ability for a first responder to utilize the map to determine which areas had been cleared, where hazards may be located, and where victims were trapped. Points were granted by the judges that reflected their feelings on this measure.

The above mentioned elements were numerically combined according to a schema that took into account merit factors that concerned (1) victims' discovery, (2) mapping, and (3) exploration. The exact point calculations for each factor are presented below.

1. 10 points were awarded for each reported victim ID. An additional 20 points were granted if the victim's status was also provided. Properly localizing the victim in the map was rewarded with an additional 10 points. At the referee's discretion, up to 20 bonus points were granted for additional information produced. For example, some teams managed to not only identify victims, but to also provide pictures taken with the robot's cameras. For this additional information teams were awarded with 15 bonus points.
2. Maps were awarded up to 50 points based on their quality, as previously described. The obtained score was then scaled by a factor ranging between 0 and 1 that measured the map's metric accuracy. This accuracy was determined through the use of the RFID scoring tags.
3. Up to 50 points were available to reward exploration efforts. Using the logged position of every robot, the total amount of explored square meters (m^2) was determined and related to the *desired* amount of explored area. This desired amount was determined by the referees and was based on the competition environment. For example, in a run where 100 m^2 were required to be explored, a team exploring 50 m^2 would receive 25 points, while a team exploring 250 m^2 would receive 50 points, i.e. performances above the required value were leveled off.

On the penalization side, 5 points were deducted for each collision between a robot and a victim. Finally, the overall score was divided by $(1+N)^2$, where N was the number of operators involved. So, completely autonomous teams, i.e. N=0, incurred no scaling, while teams with a single operator had their score divided by 4. No team used more than one operator.

It should be noted that except for the map quality, all of the above components were automatically computed from the information logged during the competition. Therefore subjective opinions during the scoring stage were reduced to the minimum. In an ideal scenario, the scoring step would be completely automatic as is currently the case for the RobocupRescue Simulation agent competition.

In addition to assigning points to determine the overall best systems, the judges assigned winning teams in the special categories of map creation and human-machine interface. The map creation award was presented to the team that consistently scored the highest in the map quality assessment while the human-machine interface award recognized the team with the most innovative robot control console.

The winning teams from the 2006 RoboCup Rescue Virtual Competition were:

First Place – Rescue Robots Freiburg, University of Freiburg, Germany

Second Place – Virtual IUB, International University Bremen, Germany

Third Place – UVA, University of Amsterdam, The Netherlands

Best Mapping – UVA, University of Amsterdam, The Netherlands

Best Human-Computer Interface – Steel, University of Pittsburgh, USA

V. COMPETITION FUTURE ASPECTS

As in all good competitions, the Rescue Virtual Competition must evolve in order to continue to challenge the competitors. In order to keep up with the competition changes, the metrics must evolve as well. While no firm decisions have been made about next year's competition, the following presents some of the current ideas.

- Victim discovery will be modified to require not only discovery, but will also require that a team provide a "data sheet" for each discovered victim. In order to receive full score, this sheet will need to include a map to the victim, information about the victim's status, and any hazards that exist along the route to the victim.
- The mapping requirement will remain the same. However, additional emphasis may be placed on including annotations on this map. Annotations will include hazards, "cleared areas", victim locations, and routes that robots took.
- Exploration will be based on the amount of area that a robot "clears". Where the definition of clearing an area means that all hazards and victims in the given area have been localized.
- Penalties will be assigned for victim bumping as well as reporting an area as clear that has victims or hazards. Inaccurate maps to victim locations will also be penalized.

VI. SUMMARY

This paper has presented results from the first annual RoboCup Rescue Virtual Competition that took place in June 2006 in Bremen Germany. The evaluation metrics for the competition were discussed, and possible modifications for the future were presented. Next year's competition will take place in Atlanta, GA. Everyone is invited to download the open source software (http://sourceforge.net/projects/usarsim/) and participate.

References

1. Asada, M. and Kaminka, G., "An Overview of RoboCup 2002 Fukuoka/Busan," *RoboCup 2002: Robot Soccer World Cup VI*, edited by G. Kaminka, P. Lima, and R. Rojas Lecture Notes in Artificial Intelligence (LNAI), Springer, 2002, pp. 1-7.

2. Carpin, S., Lewis, M., Wang, J., Balakirsky, S., and Scrapper, C., "Bridging the gap between simulation and reality in urban search and rescue," *2006 RoboCup Symposium*, 2006.

3. Carpin, S., Wang, J., Lewis, M., Birk, A., and Jacoff, A., "High fidelity tools for rescue robotics: Results and perspectives," *2005 RoboCup Symposium*, 2005.

4. Kitano, H. and Tadokoro, S.. RobocupRescue: A Grand Challenge for Multiagent and Intelligent Systems. AI Magazine 1, 39-52. 2001.
Ref Type: Magazine Article

5. Lima, P. and Custódio, L., "RoboCup 2004 Overview," *RoboCup 2004: Robot Soccer World Cup VIII*, edited by D. Nardi, M. Riedmiller, and J. Santos-Victor Lecture Notes in Artificial Intelligence (LNAI), Springer, 2004, pp. 1-17.

6. Society of Automotive Engineers, "Vehicle Dynamics Terminology," SAE, J670e, 1976.

7. Tomoiki, T., "RoboCupRescue Simulation League," *RoboCup 2002: Robot Soccer World Cup VI*, edited by G. Kaminka, P. Lima, and R. Rojas Lecture Notes in Artificial Intelligence (LNAI), Springer, 2002, pp. 477-481.