FCPx: A Tool for Evaluating Teams' Performance in RoboCupRescue Simulation League

João Certo¹, Nuno Cordeiro¹, Francisco Reinaldo^{1,2}, Luís Paulo Reis¹ and Nuno Lau³

¹ LIACC, Faculty of Engineering, University of Porto, Portugal
² GIC, Department of Computer Science, UnilesteMG, Brazil
³ IEETA-DET, University of Aveiro, Portugal
reifeup,lpreis,joao.certo{@fe.up.pt}, nuno-cordeiro@hotmail.com,
lau@det.ua.pt

Abstract. In the RoboCupRescue environment, one of the biggest problems in choosing the best team strategy is the inability to properly compare the advantages and limitations each strategy provides. This paper presents a new tool for the performance evaluation of Rescue Simulation agents, overcoming the limitations imposed by using solely the overall score. By extracting and processing a wide set of parameters from the simulation log files, and making a comparative analysis in spreadsheets, our tool facilitates the evaluation of the different agent types. By integrating our software into a widely used viewer, we expect further development to be continued by the Rescue community, allowing it to become a standard choice for the evaluation and analysis of new team strategies and approaches.

1 Introduction

RoboCup Rescue Simulation is an international joint project [1] that promotes research in distributed artificial intelligence and intelligent robotics. Through the use of a comprehensive urban disaster simulator, heterogeneous agents' teams try to minimize damage to both people and property. Every year, a RoboCup international competition is organized, where, in a competitive but constructive environment, researchers from all over the world can test their agents against other teams. Comparing approaches and exchanging ideas, progress is made at an amazing rate due to the open source nature of the project. After each competition, the source code for every team is released, so that work may be done on top of the best ideas and implementations.

The simulations take place in the three days following a major calamity in an urban scenario. Burning buildings, people trapped under debris, and blocked roads, are just some of the challenges rescue teams must overcome, coordinating as many as up to forty agents of six different types. Fire Brigade, Police Force, and Ambulance Team agents act in a coordinated manner, assisted by their

© A. Gelbukh, C.A. Reyes-García. (Eds.) Advances in Artificial Intelligence. Research in Computing Science 26, 2006, pp. 137-148 Received 02/06/06 Accepted 03/10/06 Final version 20/10/06 138

respective Control Centres, attempting to eliminate fires and save lives. Centre agents coordinate platoons in their collaborative tasks, by trading messages of world state and defining strategies. Still, platoons have the autonomy to learn, and to decide on the best course of action, on unexpected states. Police agents have the responsibility to clear roads, blocked by debris, allowing other agents to circulate and reach their targets; Ambulance Team agents search for buried citizens, rescuing and transporting them to refuges; and Fire Brigade agents extinguish buildings on fire. All tasks performed by mobile agents are centred on saving lives and reducing city damage [2][3].

The team success is measured as follows:

- 1. Let Sint be the total amount of health points of all agents at start and Bint be total undamaged area at start
- 2. At any given time step, the following values are obtained: the number of living agents (P), the remaining total health points of all agents (S) and the total undamaged area of buildings (B).
- 3. Finally, the simulation score V is calculated using Eq. 1 [4].

$$V = \left(P + \frac{S}{Sint}\right) * \sqrt{\frac{B}{Bint}} \ [1]$$

Every year, new strategies and algorithms allow successful Rescue operations on harder scenarios. A major obstacle in the development of Rescue Agent Teams is the inability to objectively measure the performance of an agent. This proves to be a big obstacle to the process of implementing minor improvements and tweaking existent parameters. The overall score, although being a single numeric value, can be affected by so many variables and starting conditions, that it becomes very difficult to describe the improvement or deterioration of the agents' performance, in each small task. Furthermore, due to the stochastic nature of the simulator, it is nearly impossible to realize if minor changes in the code improve the overall behaviour of the agents, without running several simulations using different maps. It becomes therefore essential to extract further data from the simulation. Existing tools provide a limited amount of data, while some are currently outdated and, consequently, not working. Comparison can only be presented visually, and on a limited set of parameters, with no further options to record and thoroughly analyze the observed data. It is also, usually, limited to two teams at a time, restricting the user's ability to analyze a certain aspect in a wide variety of teams.

There are two possible paths to the extraction of simulation data, which are not, in any way, mutually exclusive. The required data can be logged from the simulator or from the agents and, although some figures may be common, the different nature of both kinds of data gives each one its own perspective on the simulation. Information extracted from the agents is extremely flexible, depending, both in amount and in format, on the implementation of the code for each

individual team. Due to the lack of standardization in the Rescue community it is extremely difficult to use this data to compare different teams, as terminology and information extracted can be extremely diverse. Also, since the code is developed not only by different people, with different strategies and tactics in mind, but also in different development environments and different programming languages, the challenge of creating and maintaining a tool capable of mining the relevant data in every different team, and to present it in a comparable form, is a gigantic task - one that would require a large effort from every team involved. Still, the figures obtained from the agents are of extreme importance to complement the ones extracted from the simulator, providing the programmers with valuable insight into the behavioural process of each individual agent. It is, nonetheless, a path impossible to take if one of the main objectives is inter-team comparison, due to the above mentioned lack of standardization between teams. The chosen path is, therefore, to collect data from the simulator, benefiting from the fact that a single, common, program is used as the staging grounds for the agents' activities. This enables the gathering of information in a common form to all teams, allowing objective comparisons in any of the obtained parameters.

In order to visualize a simulation in a graphic way, a viewer tool must be used. This tool is responsible for converting the simulated information into an easily perceptible form, usually showing a plant of the city, along with the position and status of every agent during the simulation. On each run, a log containing information pertaining to the entire simulation is generated for future use. An important feature, present on most viewers, is their ability to replay simulations using those log files, allowing a detailed analysis to be performed.

In this paper we present FCPx [5], a new tool for the performance evaluation of RoboCupRescue Simulation agents. By extracting a wide set of parameters from the simulation logs and allowing a comparative analysis in spreadsheets - where the data is displayed in an organized structure - our tool facilitates the evaluation of the different agent types in various aspects. The simulation figures saved in our files are either directly acquired from the log files, or inferred from that information, using appropriate algorithms. This way, researchers are provided with enough data to allow for an easier, faster and informed decision concerning the best strategies in a number of different situations. Furthermore, this plethora of information can be used as a benchmarking environment, allowing the assessment and comparison of Rescue teams in a myriad of different factors, situations and time intervals. The data extracted may also be used to feed other software applications, from simple performance evaluators to advanced learning modules [3].

In the development of this tool two main objectives were kept in mind: flexibility and expandability. These factors are utterly important due to the constant evolution of the simulator package. By integrating our software into a widely used viewer, we expect further development to be continued by the Rescue community, allowing it to become the standard choice for the evaluation and analysis of new team strategies and approaches.

The rest of this paper is organized as follows. The next section discusses related work. In Section 3 we introduce our tool and describe it in detail. Sections 4 and 5 provide some results by presenting analysis and comparison of Rescue teams using FCPx. Finally, in Section 6 we present our conclusions and future work.

2 Related Work

Log viewers are tools used to view the evolution and result of rescue simulations. Some viewers have been written so far, enabling different viewing perspectives of the simulation, but all of them lack the functionality of a good debugging viewer [6]. Some teams have performed work in this area, but the need for a more comprehensive tool was only made more visible [6]. Also, there have been some recurring requests for good team analyzers and log viewers for some time now - both in the official RoboCupRescue mailing list as well as in other Rescue documentation.

The Freiburg team, amongst others, has developed its own viewer, releasing it to the rescue community. Freiburg's 3D viewer [7] is one of the most used by the community, second only to Morimoto's 2D viewer [8] [9], which is included in the official simulator package. Freiburg's 3D viewer tool has several built-in features. A statistic view option is available to measure and evaluate a team's performance in various aspects, such as the team's efficiency in exploring, rescuing civilians or extinguishing fires. This enables an easy comparison of two teams, since it is possible to see their actual differences, by simultaneously parsing two log files⁴. Also possible is the post-processing of simulation log files, allowing the continuous, strenuous, evaluation of time-dependant parameters.

Team Damas⁵ also made some work in this area, basing their effort on Morimoto's 2D viewer and adding code to enhance the statistic interface, making it especially useful when used in conjunction with their own "autorun" program. When both these tools were used together, several simulations would be run consecutively. At the end of each, the viewer would save the relevant statistics into a file, killing itself subsequently, thus enabling a new simulation to begin. This tool provided a limited set of statistics. It is now outdated, no longer supporting recent versions of simulator log files.

3 FCP eXtended Freiburg 3D viewer

3.1 Overview

The FCPortugal eXtended Freiburg 3D viewer (FCPx) tool was developed over the simple logging features on Freiburg's 3D viewer, extending its ability to analyze and compare different teams. A critical aspect in evaluating and comparing

⁴ http://kaspar.informatik.uni-freiburg.de/~rescue3D/index.html

⁵ http://www.damas.ift.ulaval.ca/projets/RobocupRescue/Tools.php

performance is having the information stored in a flexible, adaptable, form. If the data can be easily imported into a spreadsheet, it becomes easy to build, and analyze, tables and comparative charts. From herein, the versatility of this solution gives the programmer endless possibilities. Still, the FCPx tool provides some customizable spreadsheets to reach a few elaborate outcomes, both as a proof of concept, and as a means to help users analyze their data in a predefined way.

Our initial effort went into acquiring as much data as possible from the simulation. A large amount of parameters is directly obtained from the log file, while some other are calculated using those previously gathered. Some values are kept in both absolute and percentage form, so that users may have a relative idea of the action developments, simplifying the comparison between maps and teams. Some of the numbers are, therefore, redundant. However, the effortless comparison achieved justifies that. Also worth mentioning is the focus on expandability. It is possible to add a new parameter with little effort, besides the one required for its characterization.

The information is then saved into a file with the same base name as the log file - therefore using the "date-time-team-map" filename, but with a different extension. As was previously mentioned, versatility was the main factor when considering file formats. Keeping that goal in mind, and considering essential the ability to easily export the data into a spreadsheet, CSV (Comma Separated Values) was the chosen file type. Some other advantages arise from this choice, like the potential to reduce the amount of hard disk space occupied by log files. To properly evaluate a strategy there's a need to run dozens of simulations comparing the attained results with data from previous approaches. That generates dozens of logs, which take around 5 to 15 megabytes of disk space, each. Also, in order to examine a team's evolution through time, older logs are stocked for reference. If you add to that the need of comparison to other teams' results, therefore keeping some of their logs in store, the amount of disk space used can quickly amount to dozens, or even hundreds, of gigabytes. A CSV statistical file with our chosen parameters takes around 15 to 50 kilobytes, which is less than one hundredth of the disk space used by each log. Keeping only the statistical files, and deleting the logs, allows for an extremely larger amount of data to be stored and used, improving the ability to evaluate different teams and strategies. Still, with the deletion of the log file, the ability to reproduce the simulation is lost.

From the CSV files, the data can be effortlessly imported into a spreadsheet. This creates a space where, either through tables or charts, evolutions can be perceived, behaviors studied and conclusion drawn. By this means, the user acquires the extra flexibility necessary for the analysis, and evaluation, of minor changes in the code, and their direct, and indirect, repercussions on the agents' performance.

3.2 Operation

The open-source FCPx viewer can be fully downloaded at its official website⁶. The next section presents an experiment which was created by using FCPx Tool.

When the FCPx tool is executed with the WRITE-STATS flag set to FALSE, the program behaves in the same way as the unmodified Freiburg's 3D viewer. The only exception is in the computation of the number of civilians that die due to fire, or buried under debris, where a small bug was corrected. In contrast, when the mentioned flag is set to TRUE, CSV files are created with the additional statistical data. An extra mode was created for the fast extraction of data from previously ran simulations, allowing for extra flexibility.

But, in WRITE-STATS mode, CSV files are created with the additional statistical data, using the same base name of the log file which allows the correct correlation of the statistical files. Besides that, in "WRITE-STATS, FAST" mode, no windows are opened and, after the parsing, the viewer kills itself which allows an outside program to run the parsing, and the final score is displayed which allows fast easy access to this value. This solves another issue noted during the RoboCup 2005 competition, in which the referee would run the 2D viewer with the log file - sometimes more than once - in order to obtain the score of the competing team. The "WRITE-STATS, FAST" method takes less than 5 seconds on modern machines.

4 Evaluating different types of Agents

FCPx extracts data to a flexible organized file, presenting several opportunities for the analysis of data. As a proof of concept and as a means to show the usefulness of the tool, a comprehensive spreadsheets was created, allowing the detailed analysis of rescue teams. In this section, we will present an analysis of the FCPortugal's [3] FireBrigades' performance in the KobeHard map presented during RoboCup 2005 in Osaka. In this simulation the team results were bellow expectations, which warrants a closer study.

The Fire Brigade relevant data obtained is displayed on Table 1, Table 2 and Table 3. Table 1 presents building data. The most important figure presented is the amount of undamaged buildings, but all the other values are useful in strategic analysis. The amount of burning buildings is a good indication of whether the Fire Brigades have fires under control - looking at the displayed data we can easily see that they don't. Two important records, in tactical terms, are the number of re-ignited buildings and of those preemptively flooded that caught fire. With a perfect strategy, when a building is flooded it is done with the right amount of water. Too much will waste both water and time - extremely precious resources; while too little is even worse, allowing the building to catch fire and disrupting the prevention strategy. The same hold true for re-ignition of extinguished buildings. When a building is extinguished, the Fire Brigades should take their surroundings into account in order to decide whether they should keep

⁶ http://www.fe.up.pt/~rescue/FCPx

pouring water into the building to further lower its temperature or focus their efforts on a different building.

	Building Related Data								
				Total Number of Buildings					
	В	urning Building	js i	Flooded Unburned Buildings			740		
	Extin	guished Build	ings	Number of Times Buildings Reignited			Initial Undestroyed Area		
	Comple	etely Burned B	uildings	Perc. of Bldgs. Undamaged by Fire nor Water			14895169		
		Preem	ptively W	Number of Refuges					
							1		
Time	Burning	Extinguished	Burned	Flooded	Re-ignited	Flooded That Caught Fire	Bldgs. Undamaged by Fire nor Water		
0	N/A	N/A	NIA	N/A	0	0	N/A		
1	- 6	0	0	0	0	0	99,19%		
2	- 6	0	0	0	0	0	99,19%		
3	6	0	0	0	0	0	99,19%		
4	6	0	0	0	0	0	99,19%		
5	6	0	0	0	0	0	99,19%		
6	7	0	0	0	0	0	99,05%		
7	8	0	0	0	0	0	98,92%		
8	8	0	0	0	0	0	98,92%		
9	6	2	0	0	0	0	98.92%		
10	The state of the s								
20.									
-03	No.	Comment of the Commen			3 9 Car	0	64,73%		
290	61	2	209	0	9	0	63,24%		
300	53	2	229	0	9	0	61,62%		

Table 1. Fire Brigades - Building related data.

Table 2 exhibits the Civilian data relevant to Fire Brigades. This includes de amount of Civilians killed by fire, which is an extremely important figure, due to the high value of every live. By the end of the simulation twelve people were killed by flames, which means over twelve points were lost this way.

Agent related data is displayed on Table 3. Analyzing these numbers we should notice two important facts. Firstly, there were two Fire Brigade casualties. Rescue agents' casualties are avoidable and should never take place. In order to accomplish that, agents must be tweaked in order to ensure that only necessary risks are taken. Secondly, there are two buried Fire Brigades since the beginning of the simulation and only one is rescued. With the uncontrolled blaze present in this map, one more active Fire Brigade could prove extremely helpful.

Like tables, charts are also generated automatically in FCPx spreadsheets. In Chart 1a it can be seen that only 42% of the total score was kept and most was lost due to casualties. In chart 1b it is shown that most Civilians died due to sustained injuries from collapsed buildings, but a rather significant part (17%) were burned alive. As can be seen, using FCPx it becomes possible to analyze a single type of agent.

5 Comparing Different Teams and Strategies

The tables from the previous section provide us only rough guidelines of the team's flaws and do not make it obvious what the coding priorities should be. In

 ${\bf Table}\ {\bf 2.}\ {\rm Fire}\ {\rm Brigades}$ - Civilian related data.

 ${\bf Table~3.}$ Fire Brigades - Agent related data.

				Agent Related Data			
	Burried Po	olice Forces	Killed by Fire	FireBrigades Killed by Fire (Not Burried)			
	Burried A	mbulances K	illed by Fire	Number of Buried FireBrigades (Alive)	Total Number of FireBrigades		
	Burried F	ireBrigades I	Cilled by Fire		13		
	Buried	Agents Kille	d by Fire				
Time	Police	Ambulance	FireBrigade	FireBrigades Killed by Fire (Not Burried)	No. Buried FireBrigades (Alive)		
0		0	0	0	N/A		
1	8	0	0	0	N/A		
2		0	0	0	2		
3	8	0	0	0	2		
4	8	0	0	0	2		
5		0	0	0	2		
6		0	0	0	2		
7		0	0	0	2		
8	8	0	0	0	2		
9	8	0	0	0	2		
10		0	0		La sala di Santa 2 di La Maria di Sala		
20	Charles Sales	e	and the second				
574	A CONTRACTOR				1		
270	0	0	0	2	1		
280		0	0	2	1		
290	8	0	0	2	1		
300		0	0	2	1		

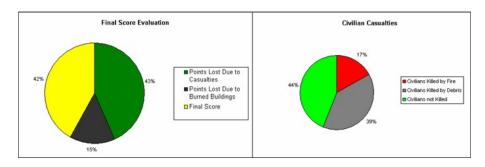


Fig. 1. a) Final score evaluation; b) Civilian casualties.

order to acquire that knowledge, we should compare our team with the current RoboCupRescue World Champions - team Impossibles, from Iran . This will allow us to perceive the most important differences. Since the competition has an open source policy, it also allows any team to analyze its weaknesses and, finding a team which performs better, adapt their strategies. For this purpose another custom spreadsheet was developed - this one oriented towards team comparison.

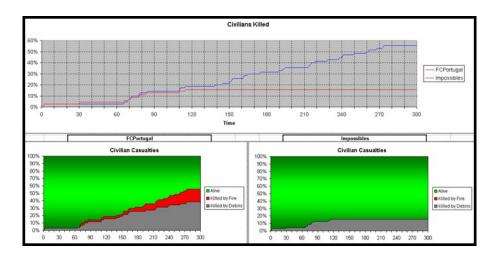


Fig. 2. a) Civilians Killed; b) FC Portugal - Civilians casualties (stacked); c) Impossibles - Civilians casualties (stacked).

The enormous difference in Civilian casualties becomes clear on Chart 2. While no Civilians were burned to death on the Impossibles' simulation, around 17% of the Civilians present in FC Portugal's simulation died by exposure to fire. The Impossibles' Ambulances were also undoubtedly more productive as the number of casualties caused by collapses was also extremely lower.

As seen on Chart 3, our opponents are able to control the fire within the first sixty simulation cycles. Besides reducing score loss from building destruction and burned civilians, this allows their Fire Brigades to pursue other tasks: namely exploration towards Civilian discovery.

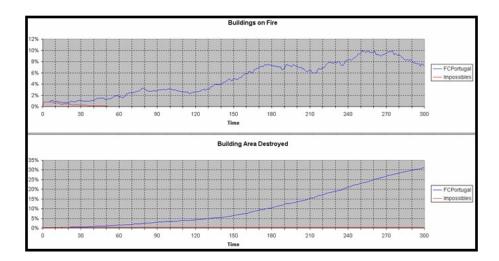


Fig. 3. a) Buildings on fire; b) Building area destroyed.

Table 4 provides us with a summary of team's comparison. Some of the facts commented on before are assembled here, allowing for some conclusions to be consolidated. The necessity to improve the Fire Brigades becomes obvious for three primary reasons: reduce the amount of points lost to burned buildings; reduce the amount of points due to burned Civilians; and allow the Fire Brigades to act as scouts when they are able to rapidly control the fires.

Display Name:	FCPortugal	Impossibles	"You may change the Display Names as you please		
Team:	FCPortugal	Impossibles			
Map:	KobeHard	KobeHard			
Date:	07 15	07 15			
Time:	10:29:20	10:58:55			
Score Data	FCPortugal	Impossibles	Civilians Discovered	FCPortugal	Impossibles
Total Number of Agents	99	99	Civilians Discovered by FireBrigades*	10	43
Number of Agents not Killed	58,59%	88,89%	Civilians Discovered by Police Forces*	44	26
Percentage of Building Area Destroyed	31,19%	0,39%	Civilians Discovered by Ambulances*	5	4
			Civilians not Discovered*	12	0
Initial Score	100,00	100,00	*Note: As explained in the civilians tab, these parts		
Points Lost Due to Civilian Casualties	43,18%	11,29%	may not add up to the total number of civilians.		
Points Lost Due to Burned Buildings	14,98%	0,19%			
Final Score	41,84%	88,51%	Civilians Discovered (Normalized by Agent Type)	FCPortugal	Impossibles
			Total Number of FireBrigades	13	13
Civilian Casualties	FCPortugal	Impossibles	Total Number of Police Forces	10	10
Total Number of Civilians	70	70	Total Number of Ambulances	6	6
Civilians Killed by Fire	17,14%	0,00%	Civilians Discovered by Each FireBrigade	0,77	3,31
Civilians Killed by Debris	38,57%	15,71%	Civilians Discovered by Each Police Force	4,40	2,60
Civilians not Killed	44.29%	84.29%	Civilians Discovered by Each Ambulance	0.83	0.67

Table 4. Comparison summary.

6 Conclusions and Future Work

The FCPx tool usefulness was proved just after its introduction in the rescue community. Several teams use it now to compare their different strategies and to compare their best strategies with other teams' strategies. Also, its open source nature and its integration into the known Freiburg's 3D viewer makes it a very easy to extend tool.

The tool enables to analyse team strategies in a very easy way and conclusions can be drawn and directions set by a simple analysis of tables and charts. What took before several simulations, detailed manual analysis and great attention, to be determined, can now be done in minutes. Using the files generated, it is a simple endeavour to create personalized spreadsheets, to evaluate a single set of parameters in different teams.

We hope our tool can solve or simplify some of the issues associated with improving a rescue team. By using a different tool to run a set of consecutive simulations it becomes possible to leave a machine running agents with different strategies, for an extended period of time, while the developer focus on different tasks. Later analysis can be done in an easy, systematic fashion, allowing faster and better informed decisions, concerning development paths.

The tool's webpage has seen quite some traffic for such a restricted community. The main page had over four hundred visits and the download page has been seen over a hundred times. It was, nevertheless, expected, since FCPx fills a gap which had been previously mentioned several times in on-line discussion.

Being the first release of this tool, only groups of agents are analyzed. Future development will be oriented into analyzing the team agents as individuals, and as such the tool will allow a greater perception off exceptional situations that led to a poorer performance.

Future work is also related with the use of our tool to enable the use of machine learning techniques in the context of (removed for blind review) team. Having a high-level configurable strategy and the capability to evaluate it, automatically, also at a high-level, using different parameters, enables us to select the best strategy for each type of map, using common machine learning techniques.

Still, even in its present form its usefulness is hard to be disputed and it had a crucial part in our achievements in RoboLudens 2006. Using FCPx's abilities to correct our weaknesses, we were able to win the competition and became the current European champions.

7 Acknowledgements

This project is a collaboration between the Universities of Aveiro and Porto. The authors wish to thank FCT/PT for the fellowship granted. This research was supported by FCT under grant # POSI/EIA/63240/2004. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect those of the sponsor.

We would also like to thank Alexander Kleiner for creating Freiburg's 3D Viewer, a great piece of software that allowed us to develop the tool focused on this paper.

References

- Hiroaki Kitano, Minoru Asada, Yasuo Kuniyoshi, Itsuki Noda, Eiichi Osawa, and Hitoshi Matsubara. Robocup: A challenge problem for ai and robotics. In RoboCup, pages 1–19, 1997.
- H. Kitano, S. Tadokor, H. Noda, I. Matsubara, T. Takhasi, A. Shinjou, and S. Shimada. Robocup-rescue: Search and rescue for large scale disasters as a domain for multi-agent research, 1999.
- 3. Francisco Reinaldo, Joao Certo, Nuno Cordeiro, Luís Paulo Reis, Rui Camacho, and Nuno Lau. Applying biological paradigms to emerge behaviour in robocup rescue team. In Carlos Bento, Amílcar Cardoso, and Gaël Dias, editors, EPIA, volume 3808 of Lecture Notes in Computer Science, pages 422–434. Springer, 2005.
- 4. H. Akin, C. Skinner, J. Habibi, T. Koto, and S. L. Casio. Robocup 2004 Rescue Simulation League Rules V1.01. The RoboCup Rescue Technical Committee, 2004.
- 5. Nuno Cordeiro, Francisco Reinaldo, Jo ao Certo, Luís Paulo Reis, and Nuno Lau. Fcpx home page. http://www.fe.up.pt/~rescue/FCPx, Feb 2006.
- Arian Team. Official arian home page. http://ce.sharif.edu/~arian/, Nov 2005. Sharif University of Technology.
- Alexander Kleiner and Moritz Gbelbecker. Rescue3d: Making rescue simulation attractive to the public public.
- 8. Takeshi Morimoto. How to Develop a RoboCupRescue Agent for RoboCupRescue Simulation System version 0. The RoboCup Rescue Technical Committee, Nov. 2002
- 9. Takeshi Morimoto. Viewer for robocuprescue simulation system. http://ne.cs.uec.ac.jp/morimoto/rescue/viewer/, 2002.