Robotic Exploration Utility for Urban Search and Rescue Tasks

Kristopher M. Thornburg and Geb W. Thomas, PhD
Mechanical and Industrial Engineering, The University of Iowa, Iowa City, IA, USA
Email: {kthornbu, gthomas}@engineering.uiowa.edu

***Abstract*—In the 2006 RoboCup Virtual Rescue Robots competition, teams from different research labs developed methods for controlling teams of mobile robots in a simulated urban search and rescue scenario. The scoring procedure used in this inaugural competition rewards participants for the number of victims found, the amount of area explored in the environment, the quality of the maps created by the robot teams and penalties participants for colliding with a victim or relying on human operators. The analysis of the strategies and scores suggests that the scoring procedure may lead teams to adopt strategies that are not consistent with the needs of a real search and rescue scenario. This paper introduces Robotic Exploration Utility as a measure of exploration quality and analyzes the results of the competition based on this measure. Individual robot contributions to the system were reviewed to account for the costs associated with adding a robot to the environment, indicating that value added per robot is an important measure that is overlooked. The analysis also revealed substantial performance variation, depending on the behavior that was being rewarded, which may indicate a lack of focus for evaluative performance measures of robotic urban search and rescue systems. The Robotic Exploration Utility metric enables the research community to focus on a performance measure which reflects the needs of the domain, while allowing task performance to be easily compared across systems.**

***Index Terms*—urban search and rescue, robotics, performance measure, robotic exploration**

I. Introduction

Using robots for urban search and rescue activities first occurred in 2001 in response to the World Trade Center disaster in New York City [1]-[3]. Since that time, there has been much interest in using robots as part of urban search and rescue teams, though much of the research has resulted in anecdotal observations [1]-[4]. There is a genuine need for quantitative and repeatable research in this area. RoboCup, in an effort to encourage research, innovation, and advancement in urban search and rescue, recently introduced a new competition called Virtual Rescue Robots, focused on developing control mechanisms for robots in a virtual setting [5], [6]. By introducing a simulation competition, the costly and time-consuming mechanical aspects of the robot are eliminated, allowing the competition to focus on robotic control. The simulation also allows for repeatable trials, quantitative data collection, and cross-institutional cooperation that did not previously exist. The competition requires an open source policy for competition algorithms, further supporting the advancement of development in robotic urban search and rescue.

This effort to use a research competition to stimulate interest and progress in a research area is a growing trend. RoboCup, in which different teams compete to develop teams of robots to compete in different leagues, is an outgrowth of a desire to rapidly advance and share results within the autonomous robotics community [5], [6]. Recently DARPA has adopted the same approach with the DARPA Grand Challenge to encourage researchers to build autonomous off-road navigation robots for a large cash prize. NASA has also experimented with offering prizes for various competitions to encourage students and researchers to focus on problems relevant to NASA's needs. These competitions are a useful and fun way to advance various research goals, but their introduction is rather recent and it is unclear how quickly and effectively the competitions will achieve the research goals of the people behind their introduction.

The simulated disaster environments and robots in the Virtual Rescue Robots competition are powered by the high fidelity Unreal Tournament 2004 game engine (Epic Games, Inc., Raleigh, N.C., USA), interfaced by an open source software package called USARSim. USARSim, originally developed at the University of Pittsburgh and now supported by NIST, allows researchers to develop realistic models of robots and control them within the Unreal Tournament 2004 architecture. Unreal Tournament contains its own environment editor that allows researchers to develop their own environments to exacting specifications. This simulation allows for repeatable trials to test individual platforms or robot teams and allows for repeatable testing of performance metrics.

Many metrics to evaluate the performance of robotic systems have been created for specific domains, including robotic urban search and rescue [7]-[10], while some metrics have been designed to be general in order to apply to a wide range of robotic systems [7]. Efforts to standardize performance metrics for domain specific robotics have taken place, but are based mostly on very specific end-user requirements [8], [9]. Example requirements include weight and volume specifications,

Corresponding author is K.M. Thornburg at kthornbu@engineering.uiowa.edu

power consumption, camera resolution, specific terrain navigation, or sensor payload. Though these are very important aspects of the system to analyze, these requirements are based on the physical system and are not easy to integrate into an overall system performance metric. System performance metrics, like amount of area explored, number of targets acquired, or number of robot collisions, are intended to measure task performance [6], [11] and are generally not based on the physical aspects of the robot. Attempts have been made to integrate many task-oriented performance measures together in order to gauge robot systems against one another, such as the scoring algorithm used in the Virtual Robot Rescue competition [5], [6]. However, occasionally these integrations are weighted in such a way that reward certain behaviors that may not be in the best interest of the stakeholders, for example, rescue workers [9], [11]. While examining these measures, we have created the Robotic Exploration Utility metric, which can be used to easily compare task performance of individual robots or robot teams. In this paper, we present Robotic Exploration Utility (REU) as a viable task performance measure and demonstrate its use based on the results from the 2006 Virtual Rescue Robots competition.

II. Methods

A. Robotic Exploration Utility

Robotic Exploration Utility (REU) is a measure designed to assess the effectiveness of a robotic system exploration or target search and is based on the idea of exploration quality, introduced by Thornburg and Thomas [11]. This measure is designed to be applied to individual robots or teams of robots and is not dependent on the type of control utilized in the system. Specifically, REU is the number of targets acquired (T) over the area explored in square meters (m^2), as described in (1).

$$REU = T / m^2 \quad (1)$$

This measure is useful in several different applications and domains, not just urban search and rescue robotics. Consider scientific robotic exploration: scientists focus on exploring the area while finding as many scientifically relevant targets as possible. The same situation is true in robotic urban search and rescue, in which the mission requires much area to be explored while finding as many survivors as possible. Even in military reconnaissance or combat missions, area explored and enemy targets acquired are the mission goals. Each domain would perhaps have different standard REU values for system evaluation purposes or for operator training benchmarks.

The main justification for the REU is the combination of two priorities of robotic exploration: the successful navigation or exploration of an area and the acquisition of targets of interest. In the case of urban search and rescue, the main priority is to identify as many victims as possible within the environment. Thus, quality of search within an area is of highest importance.

B. REU Measure Validation

To validate the REU measure, the results from the 2006 Virtual Rescue Robots competition was used. The Virtual Rescue Robots competition relied on two simulated environments relevant to a real urban search and rescue situation. The first environment, a damaged, multilevel office building, contained rubble, uneven surfaces, and flames (see Fig. 1). The second environment, a rubble-filled city street, featured uneven surfaces, flaming and overturned vehicles, and a park area with trees (see Fig. 2). Both environments, developed for the competition by NIST, were vast (several thousand square meters) and contained victims dispersed throughout in a random fashion. Each simulated victim was equipped with a radio frequency identification (RFID) tag which transmitted the victim's name and relative location when the robot was within one meter of the victim. Additional information was transmitted if the robot reached a closer distance threshold to the victim. False alarms (detecting a victim that was not present) were also possible. Additional RFID tags were dispersed throughout each environment for judging and scoring purposes.

All competition participants had access to the same robotic platforms (different sized wheeled and tracked robots) and sensors (such as sonars, cameras, and laser range finders). The robotic platforms had maximum payload specifications that required participants to carefully configure the robots with sensors. Sensor feedback was simulated as closely to real sensor feedback

Figure 1. The simulated indoor environment.

Figure 2. The simulated outdoor environment.

as possible, adding to the simulator's fidelity. The number of robots used by any participating team was not limited and communications between robots was not limited by bandwidth or other constraints.

Each trial, or run, was limited to 20 minutes, in which time each robot was to explore and map as much of the area and locate as many victims as possible. Each team started from approximately the same position and explored the same environment during each run. At the end of the 20 minute run, each team was allowed 10 minutes to compile the files necessary for scoring. The files submitted by each team for scoring a run included an image file of the map created by the robot, integrated with other robot maps if multiple robots were used in a run, a list of victims found with locations, and any additional information about the victims collected by the robot, a list of RFID tags and associated locations detected in the environment, and any images of victims recorded by the robot. Additional performance measures recorded automatically by the simulator server included the amount of area explored in square meters and the number of robot collisions with victims. The judges used this information to determine the score for each run. Table 1 describes point values given for particular aspects of each run. Equation (2) shows how the points were combined to form the final score. The point values and weights for each factor are based on the Real Robot Rescue competition to attempt to bridge between the virtual and physical competitions [10].

Six participating teams were evaluated over five runs with the competition scoring algorithm and the REU measure. The number of victims found and area explored per team for each run was recorded and analyzed based on the competition algorithm and the REU, along with the average contribution from each robot. Additionally, aggregate performance measures over the five runs was analyzed in a similar manner.

TABLE I. MERIT AND PENALIZING FACTORS WITH ASSOCIATED POINT VALUES

Merit Factors	Variable	Point value
Found victim	V	10
Victim status reported	Vs	10
Victim bonus (picture, etc.)	Vb	up to 20
Map visual quality	Mv	up to 50
Map metric quality	Mm	0 to 1
Map total		Mv*Mm
Area explored (environment dependent)	A	up to 50
Penalizing Factors		
Number human operators	N	divide total score by $(N+1)^2$
False victim identification	Vf	-5
Victim collision	Vc	-5

$$Score = \frac{[(V*10)+(Vs*10)+Vb+(Mv*Mm)+A]-[(Vf*5)+(Vc*5)]}{(N+1)^2} \qquad (2)$$

III. Results

A. Scoring Algorithm Results

Table 2 describes each team's control strategy (fully autonomous, teleoperated, or autonomy with supervisory teleoperation) and shows the aggregated score over the five runs. Only one team utilized one fully teleoperated robot, two teams utilized a combination of teleoperation and autonomous activities, and three teams were fully autonomous systems. The table is sorted from highest aggregate score to lowest.

Fig. 3 presents the competition algorithm results per run for each team. There is a general improvement trend over the course of the five runs with a wider spread of scores in the later runs.

Fig. 4 indicates the per robot contribution of scores from the scoring algorithm. These data are based on the aggregated scores divided by the number of robots utilized in the runs. There seems to be two tiers, with Red, Yellow, and Green teams scoring higher with their robots than the White, Blue, and Black teams.

TABLE II. TEAM COMPOSITION AND AGGREGATED SCORE OVER THE FIVE RUNS.

Team	Type	# Robots	Total Score
Yellow	Autonomous	8	993.09
Red	Autonomous	6	833.01
White	Autonomous	6	473.13
Black	Teleop & Auto	6	341.85
Blue	Teleop & Auto	4	207.49
Green	Teleoperated	1	139.22

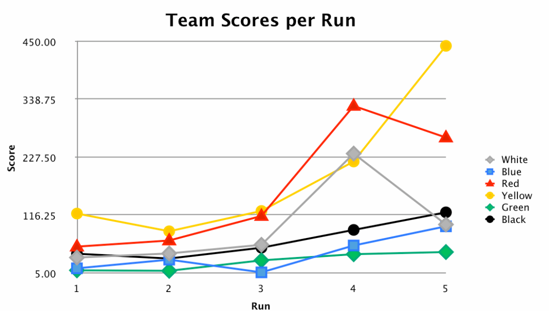

Figure 3. Competition scoring algorithm results per run for each team.

B. REU Results

Fig. 5 shows the REU for each team per run for the five runs. This figure looks significantly different from

Fig. 3, but is to be expected. Fig. 6 indicates the aggregate REU for each team over the five runs. The aggregate REU is based on the total number of victims found in the environments and the total area explored over the five runs. Fig. 7 shows the per robot contribution for the REU metric. The per robot contribution is based on the aggregate REU divided the number of robots used during the runs.

IV. Discussion

A. Scoring Algorithm

Preliminary conclusions based solely on the final scores indicate a preference to autonomous systems

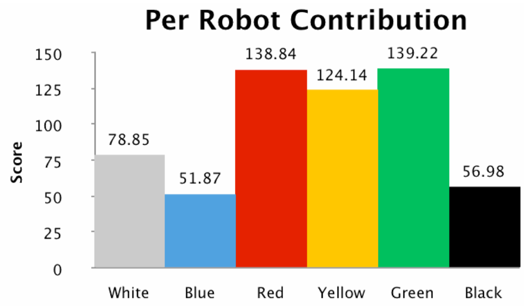

Figure 4. Per robot contribution from aggregate score.

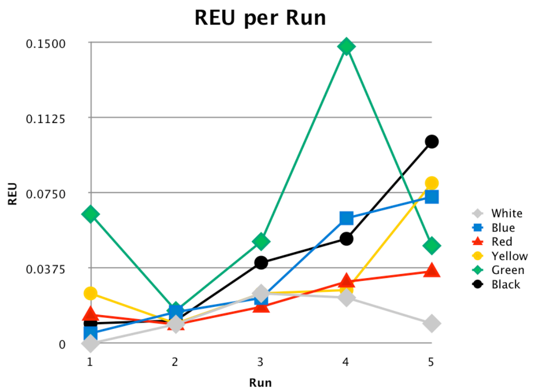

Figure 5. REU per run for each team.

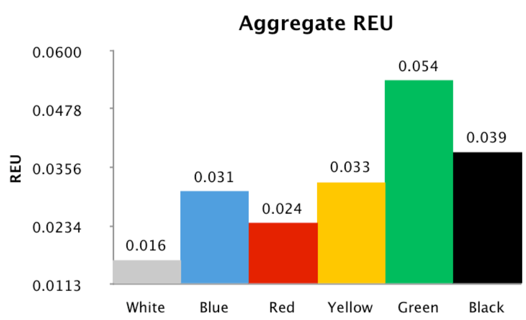

Figure 6. Aggregate REU for each team.

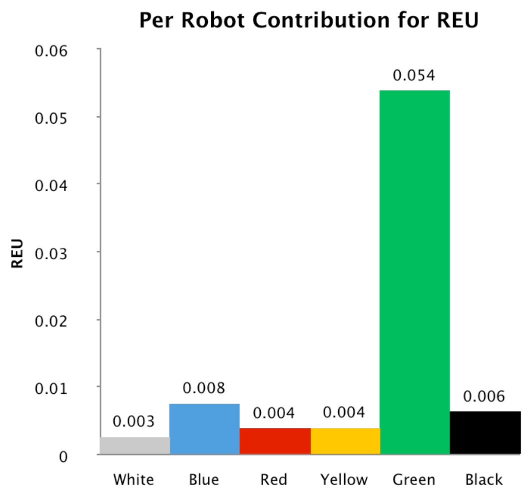

Figure 7. Per robot contribution for REU.

(see Table II) as an effective means to searching for and finding victims in an urban search and rescue environment. Table II also indicates a preference toward using more robots during a search and rescue mission, as the higher scoring teams had higher number of robots. Fig. 3 indicates the scoring algorithm results per run for each team, highlighting the fact that more robots leads to a higher score throughout individual runs as well. However, it seems that, depending on the control algorithms used, increasing the number of robots will not increase the contribution of each robot. Fig. 4 shows the per robot contribution for each team. The Green team, only using one robot, had essentially the same per robot contribution score as the Red team which was using six robots and a higher per robot contribution score than the Yellow team which used eight robots. The other three teams seemed to be relatively ineffective in using their robots.

B. REU

The REU presents an interesting measure of robot effectiveness to detect victims within the environment. Fig. 5 presents the REU per run for each team, while Fig. 6 presents the results of the aggregate calculations. The Green team has the highest aggregate REU, possibly because a human was in control of the robotic system during the entire run, rather than operating from a supervisory position like the operator of the Black team, which had the second highest REU. With the exception of the Yellow team, all the teams with a human in some sort of control capacity had a higher ratio than fully autonomous teams, indicating that the human had a direct role in locating victims within the environment. This is evident in Fig. 7, which shows the per robot contribution to the REU. The Green team, with only one robot, had a much higher per robot contribution, indicating the other teams utilized ineffective search strategies.

Different performance measures may indicate different levels of success. If the overall robotic system is considered with the competition scoring algorithm, the Yellow team created the best robotic system, as indicated in Table II. However, the aggregate REU suggests the

Green team utilized the best strategy by being far more effective in searching than the other teams. The overall score contribution per robot, as shown in Fig. 4, indicates that ineffective robot use should be a penalizing factor to reflect the cost of adding robots to the environment. The same is shown in the REU contribution per robot (Fig. 7). It is quite obvious from these differing measures that the performance measures for robotic urban search and rescue need to be standardized to focus on the most important part of a search and rescue mission.

C. Team Strategies

Each of the six teams had slightly different exploration strategies. Though 2006 was the first year for the Virtual Rescue Robots competition, five of the six team strategies have not been used in the field. The Yellow team deployed 8 robots in the environments, with the intention the robots would autonomously coordinate search efforts. Because of computation constraints, the robots independently planned each route based on the environment in a four square meter box around the robot. Each robot dropped RFID tags to ensure a collective ground truth through the team, but chose navigation paths based on their own local environment. Based on the scoring algorithm, this strategy was quite effective, as the Yellow team gained the highest aggregate score and seemed to effectively use the robots based on per robot contribution. However, the REU shows a different side, indicating that the Yellow team's exploration strategy may not have been the best (see Figs. 5 and 6). This discrepancy shows that the scoring algorithm places more weight on items that may not be fundamentally necessary to the domain.

The Red team utilized a similar strategy to the Yellow team, though only operating 6 robots in the environments. The main difference between the Red and Yellow teams is that the Red team robots attempted no information sharing during the run and each robot always acted in a selfish way. This strategy also seemed to work well given the scoring algorithm and indicates an effective use of robots based on the contribution per robot. However, the REU indicates the same effect seen with the Yellow team strategy, suggesting that the Red team exploration strategy is based more on the scoring algorithm metrics, rather than quality exploration.

The White team focused on mapping, with each of the six robots creating a separate exploration map. Each robot explored the area in a random fashion, while not sharing its internal map of the environment. The difference between the White and Red team strategies is that the White team robots navigated randomly, not selfishly. Based on the scoring algorithm, this strategy was not nearly as good as the Yellow or Red team strategies, which is supported by the contribution per robot scores. The REU shows the quality of exploration by the White team was not very good. Fig. 6 indicates this strategy is a fairly ineffectual use of robots within the environment.

The Black team used a supervisory control strategy to control six robots in the environments. The human operator had the ability to toggle modes for each robot: fully autonomous, path planning, or teleoperated. The fully autonomous mode allowed the robot to operate in a selfish manner, as used by the Red team. The path planning mode allowed the operator to assign a navigation path to a particular robot to search a particular area. The teleoperated mode allowed the operator to take full control of the robot's navigation. Based on the scoring algorithm, this strategy was not ideal. The contribution per robot score also indicates ineffectual robot exploration. However, the scoring algorithm severely penalizes teams for using a human operator (see Table I). If this penalty did not exist, the Black team would have the highest score based on the scoring algorithm [11]. The REU metric for the Black team shows a fairly effective use of robots in terms of exploration quality. Though this contradicts the results of the scoring algorithm, it is directly related to the weights applied to certain aspects of the scoring algorithm.

The Blue team strategy was similar to the Black team, though focused more on autonomous behaviors combined with teleoperation for four robots. The scoring algorithm suggests that this strategy was ineffectual compared to the other strategies. The REU metric indicates the strategy may be effective for quality exploration, as the aggregate REU is higher for the Blue team than two fully autonomous teams.

The Green team strategy was to use a teleoperated robot which carried a full payload of sensors to allow the operator to collect as much data as possible about the environment. The scoring algorithm results suggest this strategy should not be used. However, this strategy showed an effective use of the robot, as seen in Fig. 4. The aggregate REU shows the Green team strategy as the most effective for high quality exploration. The REU for each run shows the Green team being the most effective explorers four out of five runs (Fig. 5). These results imply that better exploration is achieved by keeping an operator in direct control of the robot. The Black team's use of a supervisory system is also fairly effective for quality exploration based on the REU, though some information is lost when the operator is required to attend to more than one robot. Interestingly, the Green team's teleoperated robot system is very similar to the interaction rescue workers currently have with search and rescue robots. It seems likely that, at least for the foreseeable future, urban search and rescue will be dominated by the need to use robots to search for victims in cooperation with human teams [1]-[4]. The ability of the robot to exhaustively and reliably search a region and to provide excellent documentation about what areas were searched and exactly where victims were found is likely to be the most important criteria for many years.

Like any integrated performance measure, there are limitations to the REU. The REU measure does not take into account the number of operators required to effectively operate a robot, nor does it account for the cost of adding an additional robot to the team. The REU measure also does not examine the quality of information feedback the robot or robots provide to the operator or other personnel. Additionally, the REU does not explicitly take into account the difficulty of navigating the terrain. However, the priorities of robotic exploration are reflected in the measure: area explored and targets acquired. Combining these measures creates a metric of exploration quality that can provide additional evaluative information.

V. Conclusions

Based on the results of this brief analysis, it is clear there is merit in the REU metric. A repeatable analysis of teams of robots within the USARSim structure will provide more evidence for the continued use of the REU as an assessment measure. Additionally, implementing this assessment measure can help scientists, rescue workers, and military personnel analyze and select the best systems for their purposes or facilitate the creation of training benchmarks for operators. The creation of the Robotic Exploration Utility metric is an important first step toward standardizing measures for urban search and rescue and exploration robotics in general.

References

[1] J. Casper and R. Murphy. “Human-Robot Interactions During the Robot-Assisted Urban Search and Rescue Response at the World Trade Center,” IEEE Trans. on Systems, Man, and Cybernetics Part B: Cybernetics, 33 (3), p. 367-385, 2003.

[2] R. Murphy. “Trial by Fire: Activities of the Rescue Robots at the World Trade Center from 11-21 September 2001,” IEEE Robotics and Automation Magazine, 2004.

[3] R. Murphy. "Human-Robot Interaction in Rescue Robotics," IEEE Transactions on Systems, Man, and Cybernetics - Part C, 34(2), pp. 138-153, 2004.

[4] J. Burke, R. Murphy, M. Coovert, and D. Riddle. “Moonlight in Miami: A Field Study of Human-Robot Interaction in the Context of an Urban Search and Rescue Disaster Response Training Exercise,” Human-Computer Interaction, 19, p. 85-116, 2004.

[5] M. Lewis, S. Hughes, J. Wang, M. Koes, and S. Carpin. “Validating USARsim for use in HRI research,” Human Factors and Ergonomics 49th Annual Meeting, 2005.

[6] S. Balakirsky, S. Carpin, A.Kleiner, M. Lewis, A. Visser, J. Wang and V.A. Ziparo. "Towards heterogeneous robot teams for disaster mitigation: Results and Performance Metrics from RoboCup Rescue," Journal of Field Robotics 24(11-12), pp. 943-967, 2007.

[7] A. Steinfeld, T. Fong, D. Kaber, M. Lewis, J. Scholtz, A. Schultz, and M. Goodrich. "Common Metrics for Human Robot Interaction," Proceedings of Human Robot Interaction 2006, pp. 33-40, 2006.

[8] E. Messina and A. Jacoff. "Performance Standards for Urban Search and Rescue Robots," Proceedings of the SPIE Defense and Security Symposium, April 2006.

[9] E. Messina and A. Jacoff. "Measuring Performance of Urban Search and Rescue Robots," IEEE Conference on Technologies for Homeland Security, 2007.

[10] A. Jacoff, E. Messina, B. Weiss, S. Tadokoro, and Y. Nakagawa. "Test Arenas and Performance Metrics for Urban Search and Rescue Robots," Proceedings of the 2003 International Conference on Intelligent and Robotic Systems, 2003.

[11] K. Thornburg and G. Thomas. "The Scoring Procedure for a Competitive Research Competition Influences the Usefulness of the Results in Real-World Applications," Proceedings IEEE International Conference on Systems, Man, and Cybernetics, 2007.

Kristopher M. Thornburg was born in Mount Pleasant, Iowa in 1982. He graduated from the University of Iowa, Iowa City, Iowa, with a B.S.E. in industrial engineering in 2005 and a M.S. in industrial engineering in 2006, focusing on human factors engineering.

He is currently pursuing a Doctorate degree at the University of Iowa, also in industrial engineering, with a focus on human-machine interaction.

Mr. Thornburg is a student member of IEEE, the Systems, Man, and Cybernetics Society, and the Human Factors and Ergonomics Society.

Geb W. Thomas was born in New York, New York in 1966. He received a B.S. in physics from the State University of New York, Stony Brook in 1991, as well as a Masters and Ph.D. in industrial engineering from the Pennsylvania State University in 1995 and 1996, respectively.

In 1996 he was a National Research Council Fellow at the NASA Ames Research Center, in Moffett Field, CA. In 1997 he accepted a position as Assistant Professor of Industrial Engineering at the University of Iowa, where is is now an Associate Professor and Program Coordinator for Industrial Engineering. He has authored over 70 articles on the topics of human interaction with robots and human interaction with complex systems.

Prof. Thomas is a member of IEEE, HFES, and IIE and has served as the Chair of the Virtual Environments Technical Group in HFES as ic currently a member of the Human-Systems Technical Group IEEE Systems, Man and Cybernetics and a member of the board of directors of the Computer and Information Systems Division of the IIE..