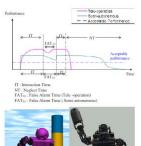


MON-AM1 Model-based Performance Evaluation Chairs: M. Ani Hsieh & Paul Evans

Validating Extended Neglect Tolerance Model with Humanoid Soccer Robots across Varying Task Complexities

R. E. Mohan, W. S. Wijesoma, C. A. A. Calderon, C. Zhou Singapore Polytechnic

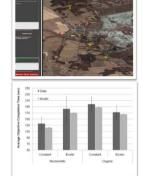
- Task complexity is a key parameter that directly impacts the robot performance as well as the false alarms occurrences.
- Validate the extended neglect tolerance model for two robot tasks of varying complexity levels.
- Experiments were performed with real and virtual humanoid soccer robots across teleoperated and semi-autonomous modes of autonomy.
- Measured false alarm demand and robot performances were largely consistent with the extended neglect tolerance model predictions.
- Experiments showed that the task complexity is directly proportional to false alarm demands and inversely proportional to robot performance.



Modeling Multiple Human Operators in the Supervisory Control of Heterogeneous Unmanned Vehicles

Brian Mekdeci, M.L. Cummings Humans & Automation Lab, Massachusetts Institute of Technology

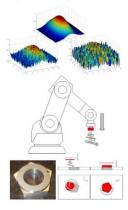
- Problems relating to futuristic teams of unmanned vehicle operators controlling multiple unmanned vehicles are discussed.
- A multiple-operator, multiple-vehicle discrete event simulation model that extends an existing single-operator model is presented.
- A design of experiment, using a video game-like a test-bed to gather empirical data, is detailed.
- Results from the experiment are compared against outputs from the model.
- Ability to model different team structures and environmental conditions is demonstrated.



Internal Model Generation for Evolutionary Acceleration of Automated Robotic Assembly Optimization

Jeremy A. Marvel and Wyatt S. Newman
Case Western Reserve University, Electrical Engineering and Computer Science Department

- A process of utilizing internal models to assist in the performance of Genetic Algorithms-based optimization strategies is discussed.
- Models are manifested as approximations of the mappings of parameter-space gene sequences to their expected (averaged) performances.
- Mappings are generated online based on empirical evaluations, and are used to predict the performances of future generations of prospective gene sequences.
- Internal models are used to optimize parameters for both robotic assembly trials and simulated problems with varying degrees of added noise.

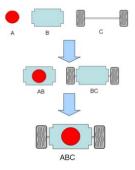


Development of Top-Down Analysis of Distributed Assembly Tasks

Anthony Cowley University of Pennsylvania

M. Ani Hsieh Drexel University C.J. Taylor University of Pennsylvania

- Describes continuous modeling techniques for discrete product assembly tasks concurrently executed by many agents
- Investigates divergence between continuous and discrete models due to agent behavior design
- Presents algorithm design approaches that allow for the continued application of continuous differential models

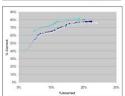


Context-Based Object Recognition

Shaun Edwards, Meredith Wright, Ben A. Abbott, Ph.D. Southwest Research Institute®

- Classical approaches to object recognition.
- · Shortcomings of classical approaches.
- Demonstrate criticality of context in human object recognition.
- Description of methods developed to integrate context into object recognition.
 - 2 Heuristic approaches
 - · 1 Mathematical approach (Bayesian)
- Discussion of the knowledge base required to support context-based object recognition.
- Measure of performance: classical approaches vs. context based object recognition.
- · Ongoing and future areas for research.





MON-AM2

Special Session I: Performance Metrics for Sustainable Manufacturing Organizers: Kevin Lyons, Mahesh Mani & Ram Sriram

Manufacturing Unit Process Life Cycle Inventories (Uplci)

Professor Michael Overcash, Wichita State University Professor Janey Twomey, Wichita State University Professor Jacqueline Isaacs, Northeastern University

- Manufacturing plants represent the transition and use of materials (metals/polymers) and chemicals into a final product
- Uplci thus links the chemical lci world to the product manufacturing world
- Goal to encourage development of other uplci using similar format, so that additive property is achieved

Conceptual Foundations of Energy Aware Manufacturing

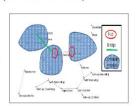
Soundar Kumara
Department of Industrial Engineering, The Pennsylvania State University

The importance of Energy Aware Manufacturing is discussed

The need for performance indices along with tasks that would be necessary to define performance indices is described

The three research components are outlined

- 1. Manufacturing Energy Computations is illustrated
- 2. Manufacturing Energy Information Modeling using web services is described
- 3. Manufacturing Energy Simulation Modeling based on Multi Agent Systems is discussed





Discrete Event Simulation as Requirements Specification for Sustainable Design of Manufacturing Systems

Björn Johansson, Anders Skoogh, Chalmers University of Technology, Sweden Mahesh Mani, Swee Leong Manufacturing System Integration Division, National Institute of Standards and Technology,

Designing manufacturing systems which lessen the impact on the environment is essential for a sustainable living. Discrete Event Simulation (DES) in combination with Lifecycle Assessment (LCA) data can be utilized to estimate how manufacturing systems will perform considering environmental measures.

- In the paper, we discuss how DES can be utilized to set requirements specifications on manufacturing systems in the early stages of the design phase.
- We present a case study of a paint shop model, that incorporates decisions on energy use, choice of machines, and environmental bottleneck detection.
 The results indicate the potential use of utilizing DES in combination with LCA data as requirements specification for designing sustainable manufacturing systems.



Figure 1. Schematic procedure for painting

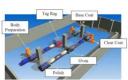
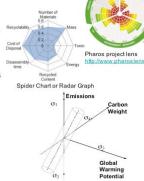


Figure 2. 3D-representation of the paint shop

Towards a New Geometric Metric for Sustainability Assessment

Gaurav Ameta School of Mechanical and Materials Engineering, Washington State University, Pullman, WA, 99164

- This paper puts forth a novel geometric metric for assessing sustainability of a product.
- The paper first reviews sustainability metrics and then focuses on the geometric metrics for evaluating sustainability aspects of a product.
- The purpose of the novel geometric metric is to present sustainability aspects of a product, for its entire life-cycle (material production, manufacturing, supply-chain, use and disposal), to design engineers in a readily comprehensible way.
- Based on the concept of areal coordinates, this paper constructs a preliminary geometric metric for sustainability assessment.



MON-PM1

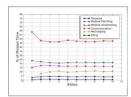
Performance Assessment and Reliability of Unmanned Systems Chairs: Hui-Min Huang & Coire Maranzano

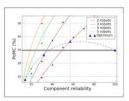
A Mission Taxonomy-Based Approach to Planetary Rover Cost-Reliability Tradeoffs

David Asikin, John M. Dolan The Robotics Institute, Carnegie Mellon University

- Motivation of using taxonomy as a framework in analyzing the reliability of planetary robotic missions is discussed.
- Architecture of the planetary robotic mission taxonomy and the criteria used to generate it are introduced.
- Environment and robot model used for simulating one of the mission classes in the taxonomy are discussed.
- Optimization of the robot configuration and tradeoff analysis between robot team size

 component reliability for Construction mission class are discussed.
- The stability of the simulation result is confirmed.

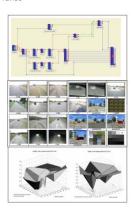




Towards a systematic assessment of the functions of unmanned autonomous systems

R. Jaulmes, E. Moliné, L.Vielle DGA, Paris Expertise Center, France

- A 4-phase generic methodology to evaluate any function of an autonomous system (UAV or UGV) is proposed.
- The method is illustrated on the evaluation of a beacon tracking algorithm for a UGV system, that was made within the TAROT study.
- The second chosen example is an evaluation of Simultaneous Localization And Mapping (SLAM) algorithms.
- Some of the relative metrics, gathered data and experimental results are presented for these two examples.
- The lessons learnt from these evaluations are described.



Performance Measures Framework for Unmanned Systems (PerMFUS)

H. Huang Intelligent Systems Division, National Institute of Standards and Technology (NIST)

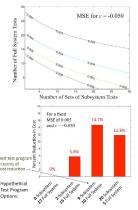
- Exploring a framework for capturing and organizing the performance metrics for Unmanned Systems (UMS)
- Based on the Autonomy Levels for Unmanned Systems (ALFUS) framework.
- Performance for UMS is to be characterizing per the missions that it is able to perform, the environments that the missions are to be performed in, the capabilities of the system itself, and the human interaction requirements.
- Axes depict increasing complexities.
 On Capability, low "levels" includes mobility mechanism types. High levels include collaboration methods.



Optimum Combination of Full System and Subsystem Tests for Estimating the Reliability of a System

Coire J. Maranzano, James C. Spall The Johns Hopkins University Applied Physics Laboratory

This paper develops a method for finding an optimum test plan, which consists of a mixture of full system and subsystem tests, to estimate the reliability of a system. An optimum test plan is developed by trading off the number of full system and subsystem tests to minimize the mean-squared error (MSE) of the maximum likelihood estimate (MLE) of system reliability and testing costs. The MSE is decomposed into the variance of the MLE and a bias from incorrectly specifying the function that relates the subsystem reliabilities to the full system reliability. Optimum test plans involve trade offs between the MSE, the degree of modeling error, and the cost of doing tests.



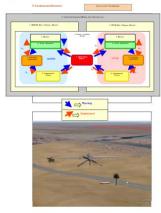
MON-PM2

Special Session II: Test and Evaluation of Unmanned and Autonomous Systems Organizers: Mauricio Castillo-Effen & Nikita Visnevski

Unmanned and Autonomous Systems Mission Based Test and Evaluation

Philipp A. Djang, Ph.D. and Frank Lopez Army Research Labs

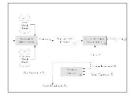
We propose to apply principles from the Army Evaluation Center's Mission Based Test and Evaluation (MBT&E) to Unmanned and Autonomous Systems (UAS) Test and Evaluation (T&E) in order to conduct rigorous, real-world testing based on anticipated military missions. In order to understand MBT&E, we introduce and describe its parent, the Mission and Means Framework. Finally, we describe a vignette that incorporates autonomous systems in the context of a mission to illustrate these principles.



Modeling and Simulation for Unmanned and Autonomous System Test and Evaluation

Mauricio Castillo-Effen, Nikita Visnevski, Raj Subbu General Electric Global Research

- Relevance of T&E in robotics
- Clear differentiation and relationship between the concepts: M&S and T&E
- Main principles of the use of M&S in T&E of robots and autonomous systems
- Metrics from mission-based capabilitydriven T&E
- Test planning as a Design of Experiments problem
- · Architecture for automated test planning
- Requirements imposed on simulation engine
- VR-Forces: A specific simulation engine chosen for test planning applied to Unmanned and Autonomous System Test & Evaluation (UAST)



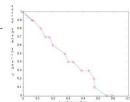


Evolutionary Framework for Test of Autonomous Systems

Raj Subbu, Nikita Visnevski General Electric Global Research Niskayuna, NY 12309 Philipp Djang Army Research Lab White Sands Missile Range, NM 88002

- DoD mission and challenge is to enable a high percentage of autonomous vehicles in the warfighter fleet by 2015.
- Autonomous systems of the future will need to be tested so their mission capabilities and robustness are predictable to the warfighter.
- Our goal is to develop an efficient intelligent test process that will enable the rapid introduction of autonomous systems on the battlefield.
- We propose a novel war game simulation-based multiobjective evolutionary test framework that combines the elements of testing an autonomous system's mission execution capabilities as a function of its innate capabilities and evolutionary computation.
- We present the evolutionary test framework, preliminary experimental results based on a limited scale war game, and ideas for developing this work into a deployable mission based test and evaluation framework.





Metrics for Co-evolving Autonomous Systems Jack Ring Educe LLC

Nominates metrics for a whole system that produces and conveys knowledge regarding Unmanned Autonomous Systems capabilities and effects.

Envisions a set of modular, distributed
Autonomous Test and Evaluation Systems
that maximize useful knowledge at
minimum investment

Envisions an Autonomous (Intelligent) Enterprise that generates customized Autonomous T&E Systems, wherever, whenever

Highlights four strategic capabilities for coevolving the whole system.

Cites proven practices and tools for achieving a unified capability.





TUE-AM1

The Role of Robotics Competitions in Advancing Intelligent Systems Chairs: Stephen Balakirsky & Jason Gorman

The Role of Robotics Competitions in Advancing Intelligent Systems: A Practitioner's Perspective

Intelligent Systems Division, National Institute of Standards and Technology (NIST)

- Competitions involving robots are becoming increasingly popular for stimulating interest and focusing efforts in engineering and science
- On one end of the spectrum, they attract young students to consider pursuing studies in science, technology and math, while at the other end, large "one time" government-funded competitions, such as the US Military's DARPA Grand Challenge serve as innovation stimulants
- In the middle, annual competitions with evolving challenges allow for continual and significant progress along a variety of technology trajectories
- This talk will discuss the role robotics competitions can play in measurement and evaluation of intelligent system capabilities and, how they can accelerate the pace of technological advancement



History and Evolution of Robot Rescue Competitions

Adam Jacoff

Intelligent Systems Division, National Institute of Standards and Technology (NIST)

Evaluating The RoboCup 2009 Virtual Robot Rescue Competition

Stephen Balakirsky¹, Stefano Carpin², Arnoud Visser³

- The format of the 2009 RoboCup Virtual Robot Rescue Competition is presented
- Details of the elemental tests and full search and rescue scenarios are presented
 - Autonomously generated map quality
 - Multi-vehicle tele-operation
 - · Communication's system deployment
- The performance metrics utilized to evaluate the competition will be presented
 - · Automatic scoring programs
 - · Hand-generated scores
- Limitations and possible improvements are discussed

¹National Institute of Standards and Technology ²University of California, Merced ³Universiteit van Amsterdam





RoboCupRescue Interleague Challenge 2009: Bridging the gap between Simulation and Reality

A. Kleiner*, C. Scrapper**, A. Jacoff***

* University of Freiburg

** The MITRE Cooperation

*** National Institute of Standards and Technology (NIST)

- The Interleague Challenge has been initiated to evaluate real-world performance of algorithms developed in simulation, as well as to drive the development of a common interface to simplify the entry of newcomer teams to the robot league.
- This paper will discuss the development of emerging test methods used to evaluate robotic-mapping, the development of a common robotic platform, and the development of a novel map evaluation methodology deployed during the RoboCupRescue competition 2009.



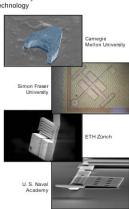




Mobile Microrobot Characterization through Performance-Based Competitions

Jason J. Gorman^a, Craig D. McGray^a, and Richard A. Allen^a
"Manufacturing Engineering Laboratory
"Electronics and Electrical Engineering Laboratory
National Institute of Standards and Technology

- A mobile microrobot competition has been organized by NIST to:
 - Assess the state of the art
 - Identify the greatest technical challenges
 - · Accelerate innovation in this field
 - Results from the past three competitions held at RoboCup will be presented along with the demonstrated technologies
 - Measuring the motion of microrobots and characterizing their performance remain challenging problems. Plans for integrating these measurements into future competitions are now being considered.



TUE-AM2

Special Session III: Is an Agent Theory of Mind Valuable for Adaptive, Intelligent Systems?

Organizer: Gary Berg-Cross

Is an Agent Theory of Mind (ToM) Valuable for Adaptive, Intelligent Systems?

Chair and Organizer: Gary Berg-Cross (Knowledge Strategies), Discussant: Alexei Samsonovich (GMU)

- Theory of Mind (ToM) propose a particular cognitive capacity by which a cognitive agent understands and predicts external behavior of others by attributing unobservable mental states, such as beliefs, desires and intentions.
- The session intends to explore the viability of the ToM concept for research and development and whether the ToM hypothesis is mature and relevant to the eventual goal of highly competent systems able to achieve goals in a relatively autonomous way.
- The session will address philosophical, research and robotic implementations and critiques central to the topic including:
 - Robotic theories of mind
 - Self-Reflecting Robots: Real Experiments with Small Societies of Minds
 - Neurodynamics of Cognition and Consciousness and Theory of Mind, Computational Tractability, and Mind Shaping





Robotic Theory of Mind

Kyung-Joong Kim Hod Lipson

Dept. of Computer Engineering Mechanical & Aerospace Engineering
Sejong University, Cornell University
South Korea Ithaca, NY 14853, USA
kimkj@sejong.ac.kr hod.lipson@cornell.edu

- Theory of mind (ToM) is a theory of cognitive function which proposes that we infer others' internal intention based on their behaviors.
- Issues include: Representation, discovery and exploitation of other's self models.
- We investigate robots representing other's self with artificial neural networks and an evolutionary learning mechanism to improve their ToM by continuous updates to weights in the network.
- Tests simulated physical robots and a novel prey-predator scenario to measure the performance of ToM learning.
- Experimental results showed that the proposed ToM approach can recover other's self models successfully.



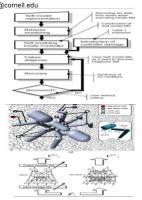


Resilient Robots

Resilient Behavior through Controller Self-Diagnosis, Adaptation and Recovery

Adaptation and Recovery
Juan Cristobal Zagal Hod Lipson
Computational Synthesis Laboratory Mechanical & Aerospace Engineering
Cornell University
Ithaca, NY 14853, USA
jcz35@cornell.edu hod.lipson@cornell.edu

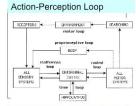
- Robot recovery behavior is investigated through a self-reflection-like process
 The select applies 2 controllers A primary.
- The robot employ 2controller: A primary "innate" reactive controller, and a secondary "reflective" controller that can observe, model and control the primary controller.
- A self-model derived observed sensorimotor time-series data data collected during normal operation of the robot is constructed by the reflective controller.
- This self-model can diagnoses the level of failure/damage introduced to the innate controller and we show that the reflective controller is able to recover performance.



Neurodynamics of Cognition and Consciousness

Robert Korzma and Department of Mathematical Sciences
The University of Memphis University of California at Berkeley

- Human cognition gathers seemingly homogeneous temporal sequences of perceptual experiences into meaningful, accessible and comprehendible chunks of fuzzy concepts
- A dynamical Theory-of-Mind (ToM) is presented to interpret experimental findings whereby meaningful knowledge is continuously created, processed, and dissipated in the form of sequences of oscillatory patterns of neural activity described through spatio-temporal phase transitions.
- The proposed approach has been implemented using intentional neurodynamics for mobile robot control in autonomous, robotic agents moving in a 2dimensional environment.

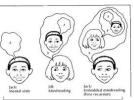




Theory of Mind, Computational Tractability, and Mind Shaping

Dr. Tadeusz Zawidzki Philosophy, Mind-Brain Institute, Mind-Brain-Evolution Cluster George Washington University

- Theory of mind = attribution of mental states like beliefs and desires.
- Problem of holism = indefinite number of distinct, finite sets of mental states compatible with any finite behavioral evidence/consequences.
- Holism makes search for correct mental state attribution potentially intractable.
- Modularity can't help any type of information is potentially relevant to "mind reading".
- Fast and frugal heuristics can't help –
 presuppose homogeneity, but good reasons to
 doubt social domain is sufficiently
 homogeneous.
- Mind shaping automatic conformism, mimicry, normative sanctioning, pedagogy make human populations sufficiently homogeneous.





TUE-PM1

Ground Truth and Testbeds for Performance Testing Chairs: Tsai Hong & Barry Bodt

Data Collection Test-bed for the Evaluation of Range Imaging Sensors for ANSI/ITSDF B56.5 Safety Standard for Guided Industrial Vehicles

Will Shackleford, Roger Bostelman NIST

- ANSI/ITSDF B56.5 Safety Standard for Guided Industrial Vehicles
- Test-Bed Hardware
- Software Architecture
 - · NML Packed Data Files
 - MOAST
 - Sensor Subsystem
 - Flash Lidar Display
 - Safety Laser Measurement Sensor
 - Camera Positioning Navigation
 - System/Spinning Laser Positioning System
- Test Results
- Changes to the Standard



Ground Truth Data Using 3D Imaging for US&R Robots

Nicholas A. Scott and Alan M. Lytle Construction Metrology and Automation Group, BFRL, NIST

- Two data collection events designed to capture ground truth data for US&R robotics research and standards development are discussed.
- Data sets were captured in 2006 and 2008 at the Disaster City training site in College Station, TX.
- Five training scenarios and two test methods captured.
- Anticipated uses:
 - Research in terrain characterization, mobility performance metrics, mapping algorithm evaluation, and developing virtual training and testing environments.
- · Data available for public use.







Performance Measurements for Evaluating Static and Dynamic Multiple Human Detection and Tracking Systems in Unstructured Environments

Barry Bodt, Richard Camden, Harry Scott, Adam Jacoff, Tsai Hong, Tommy Chang, Rick Norcross, Tony Downs, and Ann Virts

Goal:

Toward evaluation of performance for CTA detection and tracking algorithms.

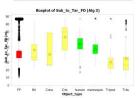
Requirements:

- · Ensure accuracy and precision of the ground truth.
- Establish time-stamp correspondences between ground-truth and detection.
- · Establish object/human correspondences
- between ground-truth and detection. · Define and compute metrics for
- performance measurements.

The assessment of algorithms include:

- Detection and misclassification rates, and false positive frequency
- · Detection distance and persistence
- · Impact of study design factors



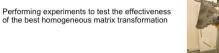


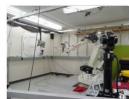
Mathematical Metrology for Evaluating a 6DOF Visual Servoing System

Mili Shah, Tommy Chang, Tsai Hong, Roger Eastman Loyola University Maryland and National Institute of Standards and Technology

Develop the best homogeneous matrix transformation to fit two streams of six degree of freedom (6DOF) data

- Challenges:
 Synchronizing two streams of 6DOF data
 - Finding a common coordinate system for two streams of 6DOF data Constructing an error metric that defines the
 - accuracy between two streams of 6DOF data







TUE-PM2

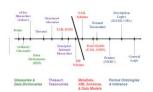
Special Session IV: An Ontology for Robotics Science and Systems **Organizers: Erwin Prassler & Herman Bruyninckx**

An Ontology of Robotics Science Herman Bruyninckx

Ontology Formalisms: What is Appropriate for **Different Applications?**

C. Schlenoff National Institute of Standards and Technology (NIST)

- We define an ontology as a knowledge representation that can be captured at different levels of formality, ranging from terms in a glossary or dictionary up to formal logic-based descriptions.
- We present and describe one way of classifying different levels of ontology formalizations
- We describe three efforts that have applied ontologies to solve real-world problems and show where they fall on the formalism spectrum.
- We show how different levels of ontology formalizations lend well to different applications.

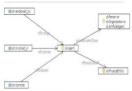


Process-Related Application A		Process-Related Application B
MATERIAL	- Work In Progress -	- WORKPIECE
STOCK -	Material	RESOURCE
RESOURCE -	Machine	MACHINE TOOL

Universal Core Semantic Layer: A Roadmap to Semantic Interoperability

Lowell Vizenor, Barry Smith National Center for Ontological Research

- The Universal Core (UCore), a central element of the US National Information Sharing Strategy, is discussed.
- How UCore Semantic Layer, a project to create an ontology-based supporting layer for UCore, is being used to further UCore's information sharing goals is
- UCore SL is an OWL DL artifact based on logically articulated definitions. UCore SL is designed to work behind the scenes in UCore 2.0 application environments as a logical supplement to the UCore messaging standard.
- UCore SL offers a logically organized vocabulary of terms, relations and definitions which can serve the semantic interoperability of UCore message content.





Mobile Robot Map-building in the Hybrid Spatial Semantic Hierarchy

Ben Kuipers

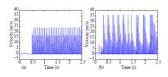
WED-AM1

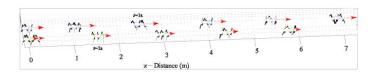
Performance Measures for Mobile Robots Chairs: Alan Bowling & Rolf Lakaemper

Performance Measures of Agility for Mobile Robots

Alan Bowling and Shih-Chien Teng The University of Texas at Arlington

- Agility involves changing velocity, which requires acceleration
- This paper examines the effect of acceleration on a known performance measure applied to a legged robot





Measuring Rover Performance in Real-time for NASA Robotic Recon Operations

D. Schreckenghost¹, T. Fong², T. Milam¹, H. Utz³

¹ TRACLabs; ² NASA Ames Research Center; ³ RIACS

- The use of real-time performance metrics for robotic reconnaissance operations during NASA's Human-Robot System field test is discussed.
- The real-time rover performance monitoring software is described.
- The performance metrics computed during the field test are defined.
- The use of these metrics during flight operations is described
 - · Metrics for flight operations
 - · Metrics for shift debrief
- The results of using these metrics in the field test are summarized.
- Conclusions for future operational use of performance metrics are presented.



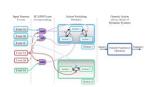


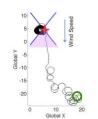


A Biologically Inspired Sensory Driven Method for Tracking Wind-Borne Odors

Brian K. Taylor, Brandon L. Rutter, Roger D. Quinn Case Western Reserve University

- Sensory Coupled Action Switching Modules (SCASM), initially based on known neural circuits in stick insects, coordinates a system to generate an emergent behavior (e.g., stepping) using sensory feedback.
- SCASM has been implemented on legged locomotion systems and provides robust, low computational cost control.
- SCASM appears to be a general control concept. Here, it is generalized and applied to the problem of odor tracking. Results are quantified with basic performance metrics.





A Confidence Measure for Segment Based Maps

Rolf Lakaemper Temple University

- Map confidence, or map quality based on regional consistency is an important measure to evaluate the quality of robot maps.
- We define a map-confidence measure that is tailored for segment based maps, without leaving the compact data representation by segments.
- The presented confidence measure is not based on comparison to ground truth data, but evaluates the map (ground truth free) based on consistency of segment clusters.
- Hierarchical clustering determines single features. Confidence measures the ambiguity in feature representation.



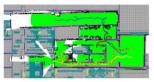


Evaluation of Robocup Maps

Ben Balaguer¹, Stefano Carpin¹, Stephen Balakirsky², Arnoud Visser³ University of California Merced¹, NIST², University of Amsterdam³

- The evolution of map scoring methods in the Robocup Virtual Rescue competition is presented
- Maps are scored based on metric quality, skeleton quality, attribution & grouping, and utility
- Examples from Robocup competitions are provided
- Steps towards a fully automated scoring procedure are presented





WED-AM2

Special Session V: TRANSTAC: Performance Evaluation of Speech Translation

Systems for Military Applications

Organizers: Craig Schlenoff & Brian Weiss

Evaluating Speech Translation Systems: Applying SCORE to TRANSTAC Technologies

C. Schlenoff, B. Weiss, M. Steves, G. Sanders, F. Proctor, A. Virts National Institute of Standards and Technology (NIST)

- The goal of the DARPA TRANSTAC program is to demonstrate capabilities to rapidly develop and field free-form, two-way translation systems.
- NIST is serving at the Independent Evaluation Team (IET) for the TRANSTAC Program.
- As the IET, NIST is responsible for analyzing the performance of the TRANSTAC systems by designing and executing multiple evaluations.
- SCORE is a unified set of criteria and software tools for defining a performance evaluation approach for complex intelligent systems.
- This document describes the TRANSTAC program and explains how the SCORE framework was applied to assess the technical and utility performance of the TRANSTAC

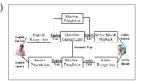




Development and Internal Evaluation of Speech-to-Speech Translation Technology at BBN

David Stallard, Rohit Prasad, Prem Natarajan BBN Technologies

- Describes BBN's speech-to-Speech (S2S) translation capability
- Compares existing and new metrics for evaluating S2S systems
- Presents a methodology for detailed analysis of translation errors and their impacts on translation quality
- Describes a novel technique for evaluating the effectiveness of audio confirmation in S2S systems
- Presents results from a systematic study of "back-translation" versus "ASR read-back" for confirmation





The Impact of Scenario Development on the Performance of Speech Translation Systems Prescribed by the SCORE Framework

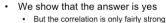
Brian A. Weiss, Craig Schlenoff National Institute of Standards and Technology

- NIST is leading the DARPA TRANSTAC evaluations to assess two-way, free form spoken language translation devices
- The SCORE (System, Component, and Operationally-Relevant Evaluation) framework is being employed to design appropriate tests to collect comprehensive quantitative and qualitative performance metrics
- High Level Concept Transfer metrics are necessary to capture technical performance of the TRANSTAC technologies
- Metrics are achieved through the development and implementation of structured and spontaneous scenarios
- These scenario types impact the metrics by the:
 - Tactical vocabulary each scenario designates Influences they have on the evaluation speakers

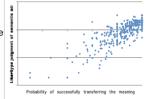
Probability of Successful Transfer of Low-level Concepts via Machine Translation: A Meta-evaluation

§ Gregory Sanders and ‡ Sherri Condon § National Institute of Standards and Technology / ‡ The MITRE Corporation

- Sentence-level judgments from bilingual judges are perhaps the most trusted metric for machine translation quality
- Would a quantitative measure of the probability of successfully transferring the meaning of the content words correlate strongly with such judgments



- We think this may reflect that fact that we gave equal weight to all content words
 - Errors are not all of equal importance



Automated Metrics for Speech Translation

Sherri Condon, Mark Arehart, Christy Doran, Dan Parvaz, John Aberdeen, Karine Megerdoomian, Beatrice Oshika, and Greg Sanders† The MITRE Corporation, † National Institute of Standards and Technology

- Automated measures of machine translation quality facilitate development of machine translation systems by providing quick and reliable feedback
- Application of these measures to translation of speech dialogues and Arabic languages has challenged the measures
- Automated metrics and their limitations are described
- Innovations were adopted to improve confidence in automated metrics for translation of speech into Arabic
- Correlations to human judgments improve with the new methods

- Ref 1: he has some pain in his stomach and always cries and complains about stomach pain
- Ref 2: he has some pain in his stomach and he always cries and says I have a stomach

Ref 3: he has some stomach pain and always cries saying my stomach hurts

Ref 4: he has a stomach ache and he always cries and says my stomach hurts

System: he has stomach pain and always crying he says pain in stomach

Utility Assessment in TRANSTAC: Using a complementary set of methods

Michelle Potts Steves and Emile Morse National Institute of Standards and Technology

Utility assessment for TRANSTAC systems is discussed. These systems are prototype realtime, speech-to-speech, bi-directional translation devices for tactical use. Utility assessment focuses on formative feedback to the research development teams in areas such as:

- Interaction factors
- System speech characteristics
- Satisfaction ratings
- · Form factor issues

Recently, the use of semi-structured interviews as a mechanism to gather data on the participants experiences using the systems has been added to the previously used survey questionnaire

How the methods were combined and the perceived impacts are discussed.





WED-PM1 Issues in Designing Intelligent Systems Chairs: Danil Prokhorov & Satyandra Gupta

Performance measurement and its role in advancement for intelligent systems: discussion points

Danil Prokhorov and Yasuo Uehara Toyota Research Institute NA, Ann Arbor, MI

- · Focus on autonomous UAV, UGV and home robots.
- · From performance measurement to performance test.
- Specifics and examples of performance test for autonomous IS.
- Common issues of testing autonomous IS:
 - · Testing vs. humans
 - · Using quantifiable performance metrics
 - · Recognizing importance of high-fidelity simulators.

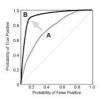
Collective Intelligence: Toward Classifying Systems of Systems

Alan J. Ramsbotham, Jr Orion Enterprises, Inc.

- Characteristics of future intelligent systems of systems are presented and discussed:
 - Ad hoc (indeterminate in terms of configuration and task).
 - Multiple mobile (or at least transportable) agents.
 - Functions associated with intelligent behavior dynamically distributed.
- Examples of different broad classes of systems are presented and possible extensions of existing conceptual frameworks to address them are postulated.
- Practical challenges to predicting performance are discussed:
 - Mission Uncertainty.
 - Inadequacy of current metrics.
 - Complexity.
 - Laws of Nature and the Myth of Information Assurance.



ALFUS Conceptua Framework

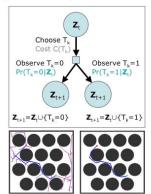


Characteristic

A Decision-Theoretic Formalism for Belief-Optimal Reasoning

Kris Hauser School of Informatics and Computing, Indiana University

- Benchmarking pitfalls: poor sampling of problem space, overfitting.
- How else can intelligent systems be optimized?
- Represent problem space using probabilistic prior knowledge
- Formalism for systems with physical and "mental" uncertainty, bounded rationality, priced information
- Formulated as a Markov Decision Process in belief space
- Belief-optimal policy solves problem optimally given prior knowledge
- Illustrated on several toy problems and a path optimization problem encountered in motion planning



A planned path (left) and smoothed (right)

Evaluation of Automatically Generated Reactive Planning Logic for Unmanned Surface Vehicles

M. Schwartz, P. Svec, A. Thakur, S. K. Gupta Energetics Technology Center, University of Maryland College Park

- Unmanned Surface Vehicles often need to utilize high speed reactive planning to carry out certain mission tasks
- We have developed a mission planning system which utilizes machine learning techniques and virtual environment to automatically synthesize a reactive planning logic for various tasks
- We have specifically used the system to develop a reactive planning logic for blocking the advancement of an intruder boat toward a valuable target
- We have developed an approach for evaluating an automatically developed decision tree representing the blocking planning logic





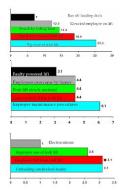
WED-PM2

Special Session VI: Performance Measurements Towards Improved Forklift Safety **Organizer: Roger Bostelman**

Fork Lift Awareness

Mark Austin Occupational Safety and Health Administration

· In this presentation, we discuss the Occupational Safety and Health Administrations' statistics on fork lift accidents and injuries caused by these accidents. Fork lift benefits and the operating environments where these manned machines work are considered. Methods used to reduce fork lift accidents as well as reported cases of lost time at work due to fork lift accidents will be discussed, along with the percentage of lift truck accident causes.



AGV Forklifts - Current and Future Safety Systems

Benny Forsman Kollmorgen Corp.

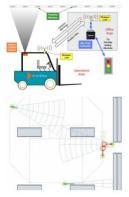
- This presentation will focus on current and future technical solutions that direct or in-direct will reduce the number of forklift related incidents. New components, new system solutions - but also new processes enabled by the new systems - will have the potential to eliminate many of the situations leading to injuries or fatalities.
- We will discuss the safety systems in AGV applications, but also the more important question of how these safe systems and practices can be used for manned forklifts and mixed fleets.



Where AGV's and Forklifts Roam: Preserving Operational Safety in a Shared Workspace

Richard H. Ungerbuehler Sky-Trax Incorporated

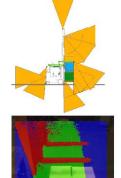
- Operational safety can be assured in facilities where industrial utility vehicles are used to transport goods only when safe practices are clearly established and carefully monitored. For example, forklift trucks, pallet jacks, buggies and carts can safely share operational space if safety is designed into the operation and vehicle operators follow safety rules and remain alert.
- In facilities where autonomous vehicles are used, a different set of safety requirements is needed. Autonomous vehicle control systems must assure that inter-vehicular collisions are prevented, and the vehicles must be equipped with safety devices, such as laser bumpers, to prevent collisions with people or equipment.



Performance Measurements Towards Improved Manufacturing Vehicle Safety

Roger Bostelman, Will Shackleford NIST Intelligent Systems Division

In this paper, we describe the current 2D (two dimensional) sensor configurations used for industrial vehicles and ideal sensor configurations for mounting 3D imagers on manufacturing vehicles in an attempt to make them safer. In a search for the ideal sensor configuration, three experiments were performed using an advanced 3D imager and a color camera. The experiments were intended to be useful to the standards community and manned and unmanned forklift and automated guided vehicle industries. The imager that was used was a 3D Flash LIDAR (Light Detection and Ranging) camera with 7 m range and rapid detection. It was selected because it shows promise for use on forklifts and other industrial vehicles. Experiments included: 1) detection of standard sized obstacles, 2) detection of obstacles while highly reflective surfaces are also within detection range, and 3) detection of forklift tines above the floor. We briefly describe these experiments and reference their detailed reports.



Group Discussion

"Recommendations for Next Generation Forklifts to Become Safer" led by: Roger Bostelman

- This group discussion will follow the session papers with a focus on directing how the next generation of forklifts can become or be made safer. Ideas to start the discussion are:
 - Better safety devices and alerts to improve driver and pedestrian awareness.
 - Better driver skills, training
 - Improved standards
 - Current or advanced technology, for example:

 - Sensors, algorithms and/or other:

 that improve situation awareness for both the forklift driver and/or pedestrians,

 - that are integrated into the vehicle controls, that can retrofit to existing forklifts, performance measurements to ensure safety and reliability.
 - Etc
- The outcome of this discussion may be put towards a white paper on the topic.



Eighteen percent of forklift accidents occur when a walking employee or other pedestrian is struck by a lift, according to Constructioninjuries.com