CPS: Frontier: VeHICaL: Verified Human Interfaces, Control, and Learning for Semi-Autonomous Systems

S. A. Seshia, R. Bajcsy, B. Hartmann, S. S. Sastry, C. Tomlin (Berkeley), R. Murray (Caltech), T. Griffiths (Princeton), C. Sturton (UNC Chapel Hill)

http://vehical.org

Award Numbers: 1545126, 1544714, 1544924
Design of Human Cyber-Physical Systems (h-CPS)

CPS that operate in concert with humans

Project Goal: To develop a science of verified co-design of controllers for semi-autonomous cyber-physical systems and interfaces between humans and cyber-physical components
Why is this Important?

SAFETY-CRITICAL & MISSION-CRITICAL

Tesla driver dies in first fatal autonomous car crash in US

IMPACT OF AUTOMATION ON WORK/JOBS

Report: Software bug led to death in Uber’s self-driving crash

Sensors detected Elaine Herzberg, but software reportedly decided to ignore her.
An Frontier-Scale Effort needs an Inter-Disciplinary Team!

- **Psychology & Cognitive Science**
  - Griffiths
  - Bajcsy

- **Human-Computer Interaction**
  - Hartmann

- **Formal Methods**
  - Seshia

- **Control Systems**
  - Murray
  - Tomlin

- **Robotics & Perception**
  - Sastry

- **Security & Privacy**
  - Sturton

S. A. Seshia
Key Envisioned Contributions to CPS Science

• Developing a Science of Co-Design of Human Interfaces and Control
  – Turning design of h-CPS from an art to a science by systematic design and verification of human interfaces

• Making Uncertainty a first-class citizen in Verification and Control
  – New algorithms and models to deal with uncertainty in CPS dynamics and CPS design

• Bridging the Schism between Model-Based Design and Data-Driven Methods
  – A new design methodology for CPS that blends data-driven learning with formal modeling and proof engines
Design for
Effective Communication between Humans and Automation
Interaction-Aware Control

Leverage human responses to estimate human internal state, and learn human model.

\[ p(u_H|x, \theta, u_R) \propto \exp(R_H(x, u_H, \theta, u_R)) \]

\[ b_{t+1}(\theta) \propto b_t(\theta) \cdot p(u_H|x_t, \theta, u_R) \]

\[ r_R(x_t, u_H, \theta, u_R) = \mathcal{H}(b_t) - \mathcal{H}(b_{t+1}) + \lambda \cdot r_{goal}(x_t, u_H, \theta, u_R) \]

\[ u_R = \operatorname{argmax}_{u_R} \mathbb{E}_\theta[R_R] \]

Goal

Info Gathering

SAMPLE RESULT

Lane Change

Nudging In

VERIFYING ROBUSTNESS

Distracted Human

Attentive Human

[Sadigh, Sastry, Seshia; RSS, IROS '16]

[Sadigh, Sastry, Seshia: CPHS '18]
Learning and Teaching (Multiple) Task Specifications

Good Communication is Crucial

Demonstrations, Natural Language

... Cost Functions, Logical Specs.

How can we hand off control reliably and intuitively?

Learning Boolean Specifications from Demonstrations

Target Specification:

Go to a yellow tile without going on a red tile. If a blue tile is stepped on, step on a brown tile before stepping on a yellow tile

[Carroll, Shah, Ho, et al., NeurIPS’19]

Booleans (logic) specifications:
- Composable
- Non-Markovian tasks
- Leverage formal methods

Humans and machines must coordinate actions and processing

On the Utility of Learning about Humans for Human-AI Coordination

[Vazquez-Chanlatte, Jha, Ho, et al., NeurIPS’18; CPHS’18]

[Carroll, Shah, Ho, et al., NeurIPS’19]
Inferring Supervisor Safe Sets for Human-Robot Teams

- Reachability analysis can compute safe sets
- Humans may perceive safety differently (from each other and from the system designer)
- Our technique can learn these individual safety preferences
- Avoid obstacles using learned safe sets to clearly communicate obstacle detection
- False positives (human thinks obstacle is undetected) are reduced

Standard Reachability Safe Set:

Human’s Perceived Safe Set:

Avoidance with Standard:  
Avoidance with Learned:

Obstacle detection is clear

False Positives reduced

Probabilistically Safe Motion Planning Around People

Use Human Models not as Ground Truth, but to Inform *Confidence in Predictions*

**Prediction:**
Confidence-aware Human Prediction w/ Boltzmann Model

\[
P(u_H^0 \mid x_H^0; \theta, \beta) \propto e^{\beta Q(x_H^0,u_H^0;\theta)}
\]

**Planning & Control:**
Fast and Safe Tracking (FaSTrack)

\[
P(\text{Crash}(x_H^t)) = \mathbb{E}_{\beta, \theta} \int_{\mathbb{X}_t(\beta)} dP(x_H^t \mid x_H^t; \beta, \theta)
\]

[Andrea Bajcsy, Sylvia Herbert, David Fridovich-Keil, Jaime Fisac, Claire Tomlin 2018]
Probabilistically Safe Motion Planning Around People

Hardware Experiment

Large Scale Simulation

A Semantic Approach to the Design of High-Assurance Learning-Based CPS
SCENIC: Scenario Description Language

- *Scenic* is a probabilistic programming language defining distributions over scenes
- Use cases: data generation, test generation, verification, debugging, design exploration, etc.

```
from gta import Car, curb, roadDirection

ego = Car

spot = OrientedPoint on visible curb
badAngle = Uniform(1.0, -1.0) * (10, 20) deg
Car left of (spot offset by -0.5 @ 0),
    facing badAngle relative to roadDirection
```

Images created with GTA-V

S. A. Seshia]
Some Applications of Scenic

- **Data Generation, (Re)-Training**
  - More controllable, interpretable
  - Improves performance significantly
  - Rare scenarios, controlled distributions, etc.

- **Debugging Failures**
  - Vary scenarios systematically
  - Explain failures of ML

- **Design Space Exploration**

Test Hypothesis: does the car model lead to a mis-detection?
VERIFAI: A Toolkit for the Design and Analysis of AI-Based Systems [CAV 2019]

https://github.com/BerkeleyLearnVerify/VerifAI

VERIFICATION
- Fuzz Testing
- Falsification

DEBUGGING
- Failure Analysis
- Retraining
- Parameter Synthesis

SYNTHESIS

System

Environment (Scenic pgm)

Specification

Semantic Feature Space

Simulator

Search

Monitor

Error Analysis

VERIFICATION

DEBUGGING

SYNTHESIS

AIRCRAFT

AUTONOMOUS DRIVING

ROBOTICS
Case Study for Temporal Logic Falsification with VerifAI: Navigation around an Accident Scenario

Lane Keeping → Lane Change with condition: d < 15
Lane change complete

Ego Car (AV) → Cones → Broken Car

d
Modeling Accident Scenario in the SCENIC Language

# Pick location for blockage randomly along curb
blockageSite = OrientedPoint on curb

# Place traffic cones
spot1 = OrientedPoint left of blockageSite by (0.3, 1)
cone1 = TrafficCone at spot1,
facing (0, 360) deg

...

# Place disabled car ahead of cones
SmallCar ahead of spot2 by (-1, 0.5) @ (4, 10),
facing (0, 360) deg

Temporal Logic Falsification
From Models to Real World: Bridging the Gap

SPEC: Specification-Aware Distance between Behaviors

SOTER: Programming Framework for Run-Time Assurance

[HSCC 2019] [DSN 2019]

Theorem: The following is an inductive invariant:

\[ \text{Mode} = \text{SC} \land s \in \varphi_{\text{safe}} \]
\[ \lor \]
\[ \text{Mode} = \text{AC} \land \text{Reach}(s,*,\Delta) \in \varphi_{\text{safe}} \]

System guarantees:

\[ \varphi_{\text{plan}} \land \varphi_{\text{mpr}} \land \varphi_{\text{out}} \]
Formal Models Key to Co-Design

Formal, Semantic and Predictive Models of Human and Environment Behavior

- Verified Control Design
- Verified Human-Machine Interaction Design
A Selection of Other Results from VeHICaL

Abstractions for Neural Network Analysis

Vibro-Acoustical Approach to Driver Interfaces

Drowsy Driver Detection

HindSight: Bicyclist Assistance Systems

*HindSight* increases the environmental awareness of cyclists by warning them of vehicles approaching from outside their visual field. A panoramic camera mounted on a bicycle helmet streams real-time, 360-degree video to a laptop running YOLOv2, a neural object detector designed for real-time use. Detected vehicles are passed through a filter bank to find the most relevant.
Industrial Impact

- Several workshops with strong industry participation
- Open-Source Tools and Datasets
  - VerifAI, Scenic, ...
  - Drowsy Driver Dataset, Visual-Acoustic Vehicle Dataset, ...
- Tools/ideas being adopted by Industry
- Working with AAA & LG on AV scenario specification and testing at GoMentum test facility
- Advice to NHTSA project on AV Test Cases
Impact on Graduate and Undergraduate Education

• Several courses impacted by VeHICaL
• Reimagining Mobility – collaboration with Ford Greenfield Labs
  – at the Jacobs Institute of Design Innovation @ Berkeley
• Academic/industry positions for graduates from VeHICaL project

[Images with names of institutions: UIUC, Stanford, UCSC, Princeton, TRI, Waymo, See-Thru: Towards Minimally Obstructive Eye-Controlled Wheelchair Interfaces]
Broader Impacts – Girls in Engineering (GiE) VeHICaL modules

• Summer program for middle-school girls at Berkeley
• VeHICaL provided instructors/mentors, funding, content
• Modules on self-driving car technology using simple Ozobot platform
VeHICaL: Verified Human Interfaces, Control, and Learning for Semi-Autonomous Systems

**Challenge:**
- Co-design human interfaces and control for human-cyber-physical systems with provable guarantees
- Apply to semi-autonomous vehicles (ground and air)

**Solution:**
- Integrate Learning, Verification and Control
- Data-Driven Resource Rational Human Modeling
- Prototype Controllers & Interfaces, Evaluate on Testbed

**Scientific Impact:**
- Developing a Science of Co-Design of Human Interfaces and Control
- Bridging Model-Based and Data-Driven Design of CPS

**Broader Impact:**
- Significantly improve safety, security, and performance of systems where humans interact closely with automation
- Involve middle/high-school and undergraduate students in VeHICaL activities

CPS Awards 1545126, 1544714, 1544924

THANK YOU!