Induction of the Human Perception-Action Hierarchy Employed in Junction-Navigation Scenarios

A. Shaukat, D. Windridge, E. Hollnagel, L. Macchi, J. Kittler

University of Surrey, Guildford, UK

{A.Shaukat, D.Windridge}@surrey.ac.uk

INTRODUCTION

Modelling human behavior in terms of its physical, logical and cognitive correlates is innately complicated. One approach to simplifying this problem is to describe human intentions in the form of a subsumptive perception-action hierarchy [2, 1]. Perception-action methodologies assume that perceptual representation depends critically upon an agent's action capability. We thus adopt the principle that an agent's set of possible perceptual transitions exists in one-to-one correlation with intentional actions at each level of the perception-action hierarchy in order to model human cognition. In particular, we seek to infer the (hierarchical) mapping existing between domain-based action protocols and human visual representations within the scenario of car driving. To this end, we adopt the Extended Control Model (ECOM) which describes human driving behavior in terms of four distinct (but simultaneous) layers of perception-action feedback [3] (Targeting, Monitoring, Regulating, Tracking), that collectively represent every aspect of driving (from navigating to car-following) as a perception-action feedback cycle. Within this model, human visual representations are determined via the correlation of attention (measured by gaze-tracking) with a highway-code-derived scene description hierarchy involving traffic lights, lanes, roads, sign junction etc. Our goal is thus to use stochastic and structural machine learning techniques to correctly determine the active intention at each level of the ECOM hierarchy when given a set of gaze, signal, control and scene-description classifier inputs. Our ultimate aim is to do this in an adaptive, unsupervised manner.

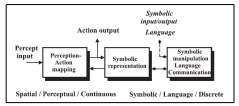

FIGURE 1: Granlund's Perception Action Model [2].

EXTENDED CONTROL MODEL (ECOM)

The Extended Control Model (ECOM) consists in the following four layers of control [3].

1. *Tracking* layer corresponds to agent's response to external disturbances, e.g. keeping a car in a specific lane or at a specific distance from the car in front.
2. *Regulating* directs tracking control layer by providing it with new goals or tasks e.g. avoiding obstacles, positioning of a car relative to other road entities.
3. *Monitoring* layer behavior in the car driving scenario tends to keep a track of all the traffic signs and signals as well as road vehicle orientation and positions.

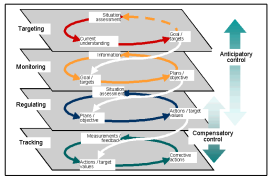

FIGURE 2: ECOM Model layers [3].

INPUT MODALITIES FOR DATA CAPTURE

Our problem is to identify *ECOM* states (i.e. hierarchical driver intentions) using the following modalities attached to an experimental car:

1. *DGPS (20Hz)*
2. *External cameras to capture traffic scene*
3. *Internal cameras for drivers eye-tracking*
4. *LIDAR (Light Detection & Ranging 20Hz sweep, 220 degrees in front of vehicle)*

GROUND TRUTH ANNOTATION

The annotation step involves the collection of suitable ground-truthed data detailing important control inputs and the driver's gaze behavior with respect to the external scene. These inputs are used to compile a comprehensive list of low-level features relating to the expertly-annotated ECOM *Regulating* and *Monitoring* level intentions and the corresponding highway-code-derived world-model. Each ECOM level consists of mutually exclusive classes that cannot be active in parallel. Gaze behavior is characterized, on a per-frame basis, using key-entity(e.g. sign, traffic-light) bounding-box transitions within both the ground-plane and view-planes. The propagation of junction topologies and zones throughout the video footage requires the aggregation of LIDAR data for approximate delineation of junction outlines, Hough-line transform and Canny edge detect to select predominating vectors. This is later fitted with a junction topology/lane structure and projected onto the screen frame. The resulting feature vector consists of 666 hierarchical binary predicates .

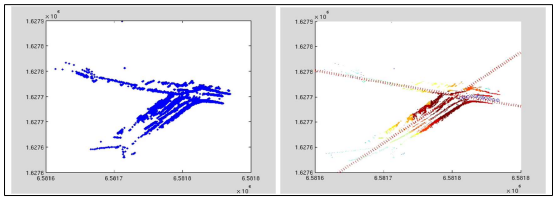

FIGURE 3: Aggregation of LIDAR data (left), edge-detected and Hough Transformed histogram (right).

CONTEXT-FREE MACHINE LEARNING

The current experiment comprises six cross-road traversing scenarios, consisting of 2 cases each of left-turning, right-turning and straight-over junction traverses. Using a *maximally populated* hierarchical domain of relational feature descriptors, ECOM-like behaviors can be learned using standard statistical pattern-recognition techniques. We here use *decision tree learners* so as to generate a discriminative but clause-based description of learning rules for later logic integration. ECOM annotations are thus split into 5-levels of hierarchical intentions per-frame constituting the class labels to be learnt. A leave-one-out cross-validation technique is used for evaluating the acontextual classifiers.

Level	Straight on 1	Straight on 2	Left turn 1	Left turn 2	Right turn 1	Right turn 2
1	0	0	0	0	0	0
2	0	0	0	0	0	0
3	10.27 ± 0.072	17.07 ± 0.15	9.10 ± 0.064	7.16 ± 0.043	29.88 ± 0.166	7.48 ± 0.046
4	15.38 ± 0.063	17.07 ± 0.15	14.13 ± 0.067	11.29 ± 0.028	30.16 ± 0.168	10.56 ± 0.035
5	16.30 ± 0.066	19.89 ± 0.079	16.81 ± 0.067	11.29 ± 0.024	25.71 ± 0.077	12.40 ± 0.041

FIGURE 4: Percentage misclassification rates for each scenario, where level (1 to 5) represent hierarchical ECOM levels of driver intentions.

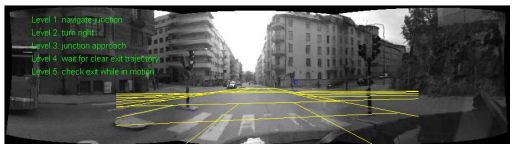

FIGURE 5: Decision tree output projected onto the screen frame along with eye-gaze (blue dot) and junction topology/lane structure.

DEDUCTIVE LOGIC SYSTEM

The logical deduction system is used as an extension to the previous acontextual intentional detection system for accommodating rule-like behaviors relating to intentional configuration changes not fully captured by stochastic correlation. In the most typical mode of operation, the logic system attempts to construct a consistent world-model from the decision-tree, computer-vision, gaze, signal and control inputs in order to determine the active ECOM intention and sub-intention at any given time. However, in the absence of decision-tree input, the logic system can also act as a per-frame ECOM intentional-classifier with performance as indicated in fig 6.

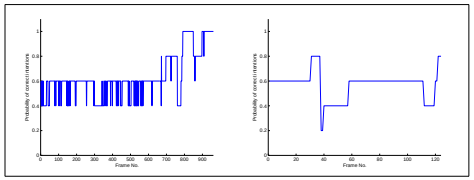

FIGURE 6: 2^{nd} example of right-turn scenario (left), 2^{nd} example of straight-over scenario (right).

It may be noticed that the accuracy figures increase with time for the right scenario, since the default 'straight on' assumption becomes falsified as more temporal context is accrued. However, by combining the decision-tree outputs with logical deduction through their incorporation into the consistency testing and aggregation procedure, the accuracy of the composite system is very significantly greater than that of the individual systems:

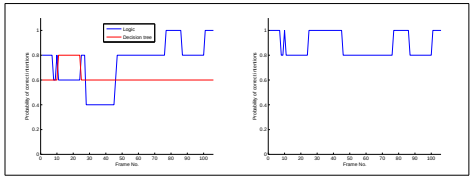

FIGURE 7: Comparison of *a priori* logic & decision-trees accuracy for 1^{st} right-turn scenario (left), combination of decision-trees with logical consistency constraints (right).

CONCLUSION

We sought to determine the mapping existing between the task-subsumption hierarchy and scene-representation hierarchy employed by humans in navigating junctions. This was accomplished via the application of decision-trees & *a priori* logical deduction to an expert-annotated hierarchy of intentional descriptors applied to driving footage with an eye-tracking overlay, as well as control and signal inputs from the car. Future work involves using logical consistency as a top-down feedback mechanism to re-weight individual detector confidences in an adaptive bootstrap cycle.

Acknowledgments: The work presented here was supported by the EU, grant DIPLECS (FP 7 ICT project no. 215078)

References

[1] Rodney A. Brooks, *A robust layered control system for a mobile robot*, Tech. report, Cambridge, MA, USA, 1985.

[2] Gösta Granlund, *A Cognitive Vision Architecture Integrating Neural Networks with Symbolic Processing*, Künstliche Intelligenz (2005), no. 2, 18–24, ISSN 0933-1875, Böttcher IT Verlag, Bremen, Germany.

[3] Erik Hollnagel and David D. Woods, *Joint cognitive systems: Foundations of cognitive systems engineering*, pp. 149–154, CRC Press, Taylor & Francis Group, 6000 Broken Sound Parkway NW, Suit 300 Boca Raton, FL 33487-2742, Feb 2005.