On Control Transitions in Autonomous Driving: A Framework and Analysis for Characterizing Scene Complexity

Nachiket Deo, Nasha Meoli, Akshay Rangesh and Mohan M. Trivedi
Laboratory for Intelligent & Safe Automobiles, UC San Diego
{ndeo, nmeoli, arangesh, mtrivedi}@ucsd.edu

Abstract

'Take-overs' are safety critical events in conditionally autonomous vehicles. These are cases where vehicle control is transferred from the autonomous system to a human driver during failure modes of the system. Safe take-overs depend on two key factors; the readiness of the driver, and the complexity of the scene. While prior work has addressed driver readiness estimation, scene complexity estimation for control transitions remains an unexplored topic. In this paper, we focus on characterizing the complexity of driving scenes as perceived by human drivers during takeover events. To this end, we collect naturalistic driving data using a conditionally autonomous vehicle, equipped with cameras and LiDAR sensors. We mine a diverse set of scenarios using the LiDAR point cloud statistics. We then collect take-over complexity ratings in these scenarios assigned by raters with varying degrees of driving experience. We present an analysis of inter-rater agreement, and the average rated complexity conditioned on features of the surrounding environment, detected agents around the ego-vehicle, and ego-vehicle actions and motion states.

1. Introduction

Conditionally autonomous vehicles are now commercially available. Such vehicles can operate autonomously under certain traffic constraints, but require a human driver to serve as backup during failure modes. To ensure safety of its occupants and surrounding traffic actors during failure modes, a conditionally autonomous vehicle can employ one of two courses of action, detailed in Figure 1. The first option would be to engage active safety features such as tightening seat-belts, safely exiting the driveable area while avoiding collisions, or deploying airbags. The second option would be to transfer control of the vehicle to the human driver, termed a 'take-over'. Take-overs can be far less disruptive for the vehicle's occupants and the surrounding traffic, if performed safely. Safe take-overs would hinge on

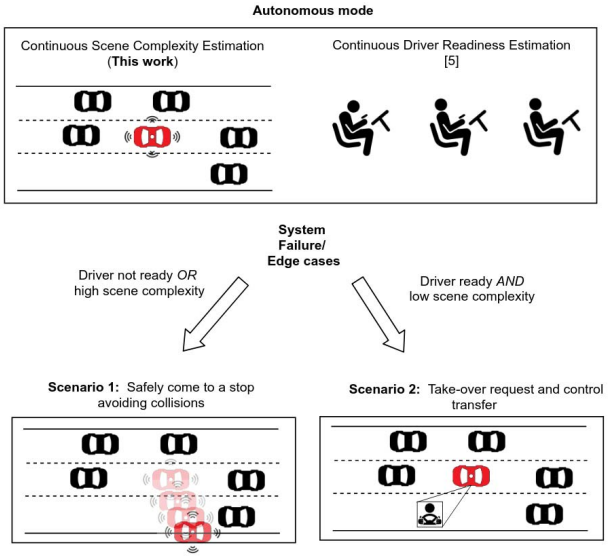

Figure 1: **Motivation**: Continuous estimation of the driver's take-over readiness and surrounding scene complexity (top). During failure modes, if the surrounding scene complexity is low and the driver is prepared to take over, control can be transferred to the driver via a take-over request (scenario 2, bottom right). If the scene complexity is high or the driver isn't prepared, automatic collision avoidance and active safety measures can be employed (scenario 1, bottom left). In this work, we focus on characterizing scene complexity from the human driver's perspective.

two factors; the preparedness of the driver to take over control and the complexity of the surrounding scene. Thus, a conditionally autonomous vehicle needs the ability to continuously estimate driver readiness and surrounding scene complexity for take-overs.

A number of studies have addressed control transitions in conditionally autonomous vehicles [1, 2, 8, 9, 10, 4, 6, 11, 3]. In particular, recent work has addressed driver readiness estimation for take-overs [2, 11, 3]. However, scene complexity for take-overs has not been extensively investigated. Currently, scene complexity is defined in terms of traffic

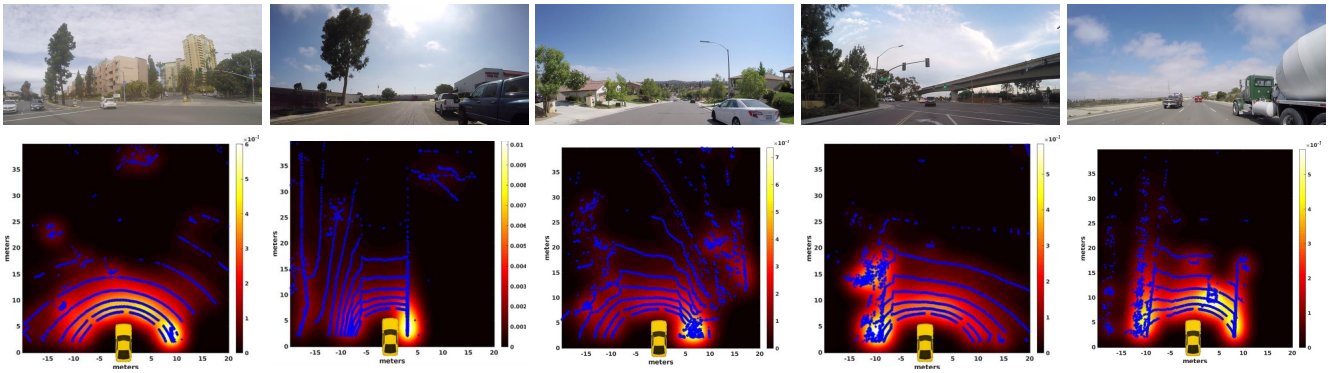

Figure 2: **Examples of diverse scenes generated by our mining algorithm:** Top row shows images from our camera, and the bottom row depicts the corresponding probability distribution (heatmap) with superimposed LiDAR points (in blue).

density as in [10, 4, 6], weather conditions, or road hazards as in [9, 8]. However, a wider range of factors might affect the take-over complexity of a scene. While there may be significant overlap in the definitions, the system boundaries of the autonomous vehicle are typically well understood by its designers, manufacturers and even end users, but the limits of a human driver are more ambiguous. In this study we seek to understand scene complexity from a human driver's perspective.

In particular, we collect naturalistic driving data using a conditionally autonomous vehicle, equipped with cameras and LiDAR sensors. We mine a diverse set of scenarios using the LiDAR point cloud statistics. We then collect take-over complexity ratings in these scenarios assigned by raters with varying degrees of driving experience. We present an analysis of inter-rater agreement, and correlation of rated complexity with features based on the surrounding environment, detected agents around the ego-vehicle and ego-vehicle actions and motion states.

2. Data Collection and Diverse Event Mining

We use the vehicle test bed described in [11] for collecting data. The test bed is built on top of a Tesla Model S, capable of driving in autonomous mode on highways and certain urban environments. The test bed is equipped with 6 high resolution cameras synchronously capturing the vehicle's complete surroundings at 30 fps. Additionally, a 16 layer Velodyne LiDAR captures range data around the vehicle. The complete dataset contains over 150 hours of naturalistic driving data captured on Californian freeways and urban streets.

Our goal is to define a metric for take-over complexity in scenes based on subjective ratings provided by human observers. The complete dataset is too large to be rated by multiple human raters. In order to generate a diverse subset of the data to be rated, we propose an automated approach for mining diverse events.

We use LiDAR point cloud maps for mining diverse events. We estimate the probability distribution of the LiDAR points projected in the birds eye view. We use bi-variate Kernel Density Estimation (KDE) at each point, given by:

$$\hat{f}_H(x) = \frac{1}{n}\sum_{i=1}^{n} K_H(\mathbf{x} - \mathbf{x_i}), \tag{1}$$

where,

- $\mathbf{x_i}$ are the LiDAR points in the bird's eye view.
- K is the bi-variate normal kernel density function

$$K_H(\mathbf{x}) = \frac{1}{2\pi\sqrt{|\mathbf{H}|}} e^{-\frac{1}{2}\mathbf{x}^T\mathbf{H}^{-1}\mathbf{x}}. \tag{2}$$

- $\mathbf{H}$ is the 2×2 bandwidth (smoothing) matrix given by $\sigma^2\mathbf{I}$, where σ^2 is experimentally determined.

This estimate is then smoothed to produce a bi-variate probability distribution representative of objects in a driving scene. Figure 2 shows examples of the smoothed probability distributions.

We select a diverse set of frames from the dataset using the Kullback-Leibler divergence (seen in Equation 3) of their estimated probability distributions. We first sample the complete dataset for frames at 5 second intervals and randomize the order of the sampled frames. We then iterate over these sampled and randomized frames to generate a subset of selected frames. A frame is added to this subset if its K-L divergence with respect to existing selected frames is greater than a threshold.

$$D_{KL}(P||Q) = -\sum_{x\in X} P(x) log(\frac{Q(x)}{P(x)}) \tag{3}$$

A total of 3128 diverse frames are selected using this process. We consider 2 second video clips centered at the

Figure 3: **Rating tool:** Screenshot from the tool used for rating of 2s video clips.

selected frames to obtain a diverse set of scenarios to be rated for scene complexity.

3. Scene Complexity Ratings

We chose a pool of 6 human raters with driving experience and working knowledge of the Tesla autopilot system to assign subjective ratings to the selected video clips. The raters were shown video feed from the forward facing camera. Figure 3 shows the interface used for collecting ratings. The raters were given the following prompt, *" You'll now be shown video clips of a vehicle operating in autonomous mode. Rate on a scale of 1 to 5 the difficulty involved in taking over control from the vehicle at the end of each clip."*. We chose a discrete rating scale rather than a continuous scale to minimize rater confusion. Each rater rated a *common set* of 90 video clips. We use the common set to normalize for rater bias and to analyze rater agreement. The remaining 3038 video clips, termed the *expansion set*, were divided across the raters. We use the expansion set to analyze the effect of cues from the surrounding scene and ego-vehicle on the assigned scene complexity rating.

3.1. Normalizing for rater bias

One source of noise in the assigned ratings is rater bias. Raters can be strict or lax, and can use a varying range of values. We normalize for rater bias using a percentile based approach. We use the common set for normalization of the ratings. We pool and sort ratings provided by each rater on the common set to obtain rater specific look-up tables. We then pool and sort ratings of all raters to obtain a combined look-up table. To normalize a specific raters ratings, we find the percentile range of the assigned value in the raters lookup table. We then replace it with the average of all values in that percentile range in the combined look-up table. This percentile based lookup can be applied to the entire dataset, including the expansion set.

Table 1: ICC values for annotator ratings

Normalization	ICC(C, 1)	ICC(A, 1)	ICC(A, k)
✗	0.522	0.232	0.644
✓	0.517	0.520	0.866

4. Analysis of Ratings

4.1. Inter-rater agreement

We use intra-class correlation co-efficients (ICCs) as formulated by McGraw *et. al.* [7], to evaluate inter-rater agreement. We model the human ratings as a two-way random-effect model without interaction, assuming n observations and k raters. Under this model, the rating x_{ij} assigned by rater j to clip i can be expanded as,

$$x_{ij} = \mu + r_i + c_j + e_{ij}, \tag{4}$$

where, μ is the global average rating, r_i's are the deviations based on the content of the rated clips, and c_j's are the deviations due to rater bias. The r_i's and c_j's are independent, with mean 0 and variance σ_r^2 and σ_c^2 respectively. And finally, e_{ij} is the normally distributed measurement error with zero mean and variance σ_e^2. We report the following ICC values for the normalized and unnormalized ratings, as defined in [7]:

$$ICC(C,1) = \frac{\sigma_r^2}{\sigma_r^2 + \sigma_e^2}, \tag{5}$$

ICC(C,1) can be interpreted as the degree of consistency of the rating values. This is independent of the rater bias, and has a high value if the trend of ratings across raters is consistent.

$$ICC(A,1) = \frac{\sigma_r^2}{\sigma_r^2 + \sigma_c^2 + \sigma_e^2}. \tag{6}$$

ICC(A,1) is the degree of absolute agreement of rater values, and has a high value only if the raters are in agreement in terms of the actual value of the ratings.

$$ICC(A,k) = \frac{\sigma_r^2}{\sigma_r^2 + \frac{\sigma_c^2 + \sigma_e^2}{k}}. \tag{7}$$

ICC(A,k) can be interpreted as the reliability of the average rating provided by k different raters. In our case, $k = 6$.

All ICC values are bounded between 0 and 1. The σ values are estimated using two-way analysis of variances (ANOVA). Koo and Li [5] prescribe that ICC values less than 0.5, between 0.5 and 0.75, between 0.75 and 0.9, and greater than 0.90 are indicative of poor, moderate, good, and excellent reliability, respectively. Table 1 shows the ICC values with and without normalization. As expected, the

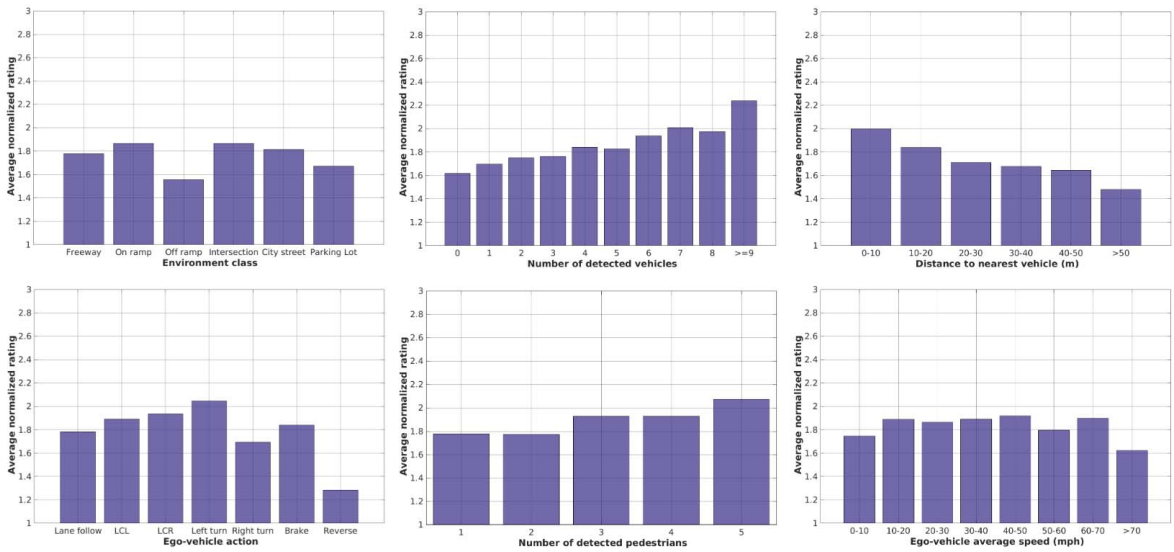

Figure 4: **Effect of environment and ego-vehicle cues on assigned ratings:** Variation in average ratings conditioned on environment class (top-left), ego-vehicle action (bottom-left), traffic density (top-middle and bottom-middle), distance to the nearest vehicle (top-right), and ego-vehicle speed (bottom-right)

ICC(C,1) values are higher than the ICC(A,1) values due to the rater bias term σ_c^2 in the denominator for ICC(A,1). However, we note that normalization considerably improves the ICC(A,1) values without affecting the ICC(C,1) values. This shows that the normalization maintains the trend (σ_r^2) of the ratings while reducing rater bias (σ_c^2). Finally, the last column shows the ICC(A,k) values, which represent the reliability of the average rating provided by all raters. ICC(A,k) also considerably improves after normalization.

4.2. Effect of environment and ego-vehicle cues on ratings

We analyze the effect of cues from the ego-vehicle and its environment on the assigned ratings. In particular, we consider the following features:

Environment class: We bin the data into 6 different environment classes shown in Figure 4 (top-left). The environment class for each video clip is manually annotated. We note that environment classes involving impending decisions or maneuvers, such as on ramps and intersections, have high ratings. Environment classes having fewer impending decisions and maneuvers such as parking lots and off ramps have low ratings.

Ego-vehicle action: We also manually annotate the ego vehicle's action for each video clip. We consider lane keeping, left and right lane changes, left and right turns, braking and reversing. Figure 4 (bottom-left) shows the average ratings conditioned on ego-vehicle action. We note that lane changes and left turns have high complexity ratings, since these maneuvers often involve interaction of surrounding traffic agents. On the other hand right turns and reversing correspond to low ratings.

Traffic density: We detect vehicles and pedestrians in front of the vehicle using the monocular vision based 3-D detector described in [12]. We note from Figure 4 (top-middle and bottom-middle), that a higher number of detected vehicles and pedestrians correspond to a higher rating.

Distance to the nearest vehicle: We also observe a trend in the average rating with respect to distance to the nearest vehicle, with lower distances corresponding to higher ratings, seen in Figure 4 (top-right). This seems reasonable, due to the higher collision risk with other vehicles close by.

Ego-vehicle speed: Finally, somewhat surprisingly, we note that the ego-vehicle speed does not seem to affect the assigned rating as seen in Figure 4 (bottom-right). This might be due to the raters underestimating the speed of the vehicle while observing the videos, or due to lower traffic densities when the ego-vehicle is moving at high speeds.

5. Concluding Remarks

We presented an approach to characterize scene complexity for control transitions in autonomous vehicles using subjective ratings provided by human raters viewing camera feed from a conditionally autonomous vehicle. We analyzed the agreement across raters in terms of intra-class correlation coefficients. Finally, we analyzed the variation in average ratings with cues from the ego-vehicle and its environment. Future work would focus on utilizing the normalized ratings as the ground truth for training machine learning models to estimate scene complexity for take-overs.

References

[1] C. Braunagel, E. Kasneci, W. Stolzmann, and W. Rosenstiel. Driver-activity recognition in the context of conditionally autonomous driving. In *2015 IEEE 18th International Conference on Intelligent Transportation Systems*, pages 1652–1657. IEEE, 2015. 1

[2] C. Braunagel, W. Rosenstiel, and E. Kasneci. Ready for take-over? a new driver assistance system for an automated classification of driver take-over readiness. *IEEE Intelligent Transportation Systems Magazine*, 9(4):10–22, 2017. 1

[3] N. Deo and M. M. Trivedi. Looking at the driver/rider in autonomous vehicles to predict take-over readiness. *arXiv preprint arXiv:1811.06047*, 2018. 1

[4] C. Gold, M. Körber, D. Lechner, and K. Bengler. Taking over control from highly automated vehicles in complex traffic situations: the role of traffic density. *Human factors*, 58(4):642–652, 2016. 1, 2

[5] T. K. Koo and M. Y. Li. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of chiropractic medicine*, 15(2):155–163, 2016. 3

[6] H. Marciano and Y. Yeshurun. Perceptual load in different regions of the visual scene and its relevance for driving. *Human factors*, 57(4):701–716, 2015. 1, 2

[7] K. O. McGraw and S. P. Wong. Forming inferences about some intraclass correlation coefficients. *Psychological methods*, 1(1):30, 1996. 3

[8] B. Mok, M. Johns, K. J. Lee, D. Miller, D. Sirkin, P. Ive, and W. Ju. Emergency, automation off: Unstructured transition timing for distracted drivers of automated vehicles. In *2015 IEEE 18th International Conference on Intelligent Transportation Systems*, pages 2458–2464. IEEE, 2015. 1, 2

[9] B. Mok, M. Johns, D. Miller, and W. Ju. Tunneled in: Drivers with active secondary tasks need more time to transition from automation. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 2840–2844. ACM, 2017. 1, 2

[10] J. Radlmayr, C. Gold, L. Lorenz, M. Farid, and K. Bengler. How traffic situations and non-driving related tasks affect the take-over quality in highly automated driving. In *Proceedings of the human factors and ergonomics society annual meeting*, volume 58, pages 2063–2067. Sage Publications Sage CA: Los Angeles, CA, 2014. 1, 2

[11] A. Rangesh, N. Deo, K. Yuen, K. Pirozhenko, P. Gunaratne, H. Toyoda, and M. M. Trivedi. Exploring the situational awareness of humans inside autonomous vehicles. In *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, pages 190–197. IEEE, 2018. 1, 2

[12] A. Rangesh and M. M. Trivedi. Ground plane polling for 6dof pose estimation of objects on the road. *arXiv preprint arXiv:1811.06666*, 2018. 4