

Video Analytics for Quantifying Driver Distraction and Engagement

Brandon M. Smith

Research Associate

Department of Computer Science

University of Wisconsin-Madison

Email: bmsmith@cs.wisc.edu

Charles Dyer

Professor

Department of Computer Science

University of Wisconsin-Madison

Office: (608) 262-1965, Email: dyer@cs.wisc.edu

Madhav Chitturi

Associate Researcher

TOPS Laboratory, Department of Civil and Environment Engineering

University of Wisconsin-Madison

Office: (608) 890-2439, Email: mchitturi@wisc.edu

John D. Lee

Professor

Department of Industrial Engineering

University of Wisconsin-Madison

Office: (608) 890-3168, Email: jdlee@enr.wisc.edu

Driver distraction represents a major safety problem in the U.S. With greater levels of automation and adoption of advanced driver assistance systems (ADAS), driver distraction/engagement becomes all the more critical. In this research, our goal is to automatically quantify driver behavior, specifically distraction and engagement. Toward this goal, we have developed an efficient computer vision system to extract high-level features associated with driver behavior from videos. We present a novel algorithm for estimating facial landmark coordinates (e.g., nose tip, eye and mouth corners). The foundation of our approach is cascaded shape regression (CSR), which has emerged recently as the leading strategy for facial landmark estimation. We propose a generalization of conventional CSRs that we call branching cascaded regression (BCR). Conventional CSRs are single-track; that is, they progress from one cascade level to the next in a straight line, with each regressor attempting to fit the entire dataset. We instead split the regression problem into two or more simpler ones after each cascade level. Intuitively, each regressor can then operate on a simpler objective function, which is especially important for handling large head pose variation. On standard computer vision benchmarks, the algorithm produces results with state-of-the-art accuracy. The system uses the landmark estimates to compute head pose (yaw, tilt, and roll angles), eye and mouth state (relative openness), and pose-related landmark visibility. Key challenges include low resolution, low dynamic range, and harsh illumination conditions, among others. Therefore, one of the key features of the system is a confidence score associated with each estimate to identify where manual involvement might be necessary. The system runs at 250+ frames per second on a modern laptop. This algorithm can be used to process the millions of hours of SHRP2 naturalistic

driving data to understand the situations where distraction and disengagement threaten driving safety and how automated vehicle technology might mitigate those risks.