The In-Car Sign Language Corpus (ICSL): A Multi-Modal Resource For Constrained-Space Sign Language Recognition

R. Boddu*, G. V. Leite†, J. L. da Silva*, A. Benetti†, I. Barbieri†, N. de M. Afonso†, T. Santos‡, H. Pedrini†, F. V. Barbosa‡, J. M. De Martino†, M. Georges*, A. Zimmer*
*THI (Germany), †UNICAMP (Brazil), ‡USP (Brazil)

Motivation (Constrained SLR)

- Seatbelts create severe physical body occlusions
- Cramped interiors restrict natural signing volume
- Dynamic lighting generates high visual noise
- Non-frontal perspectives degrade tracking stability
- Laboratory datasets lack realistic vehicle constraints

Current SLR systems often fail to generalize to in-the-wild vehicle interiors. This work establish technical baselines for automotive SLR

ICSL Corpus Overview

- ICSL is driven by the need for accessible passenger-driver communication within shared mobility services focusing on constrained signing space of vehicle interior
- Specialized multimodal resource documenting to date 127 essential phrases (defined 1,344) in context specific scenarios
- Captured manual signs and non-manual markers and developed in collaboration with Deaf researchers

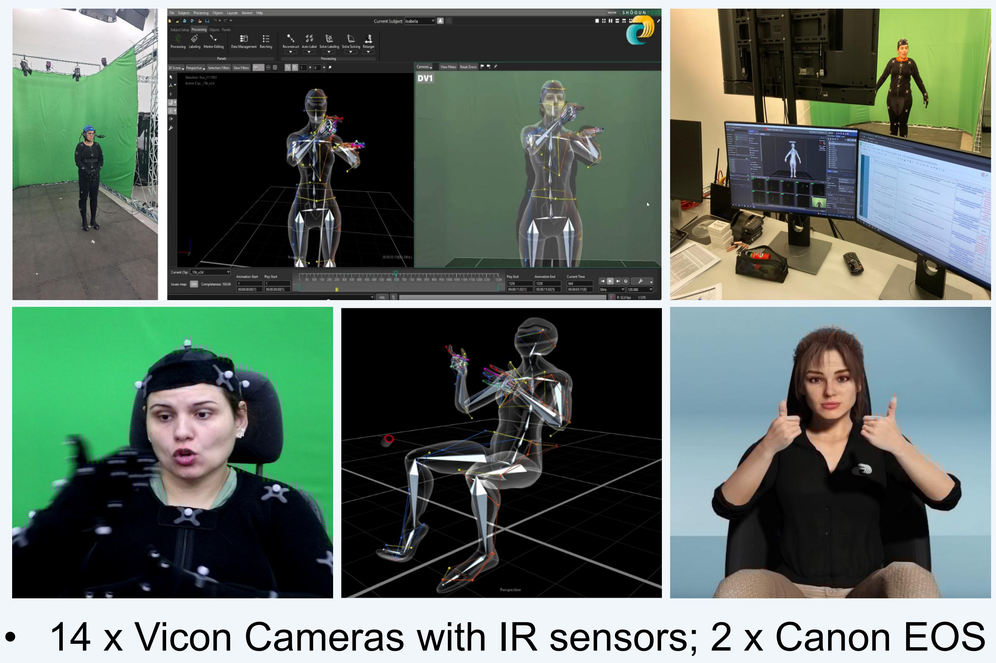
1.5M+ Frames

127* Contextual Phrases

3 Native Signers

Methodology (Multi-modal Acquisition)

Laboratory Mocap Baseline

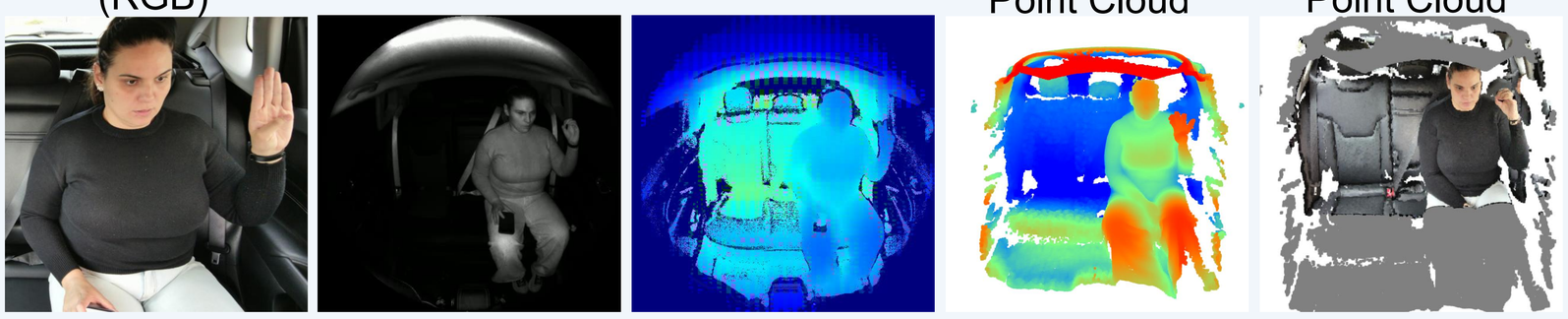


- 14 x Vicon Cameras with IR sensors; 2 x Canon EOS T6i Cameras captures RGB; 1 x FLIR 2D; 1 x Vicon Vue Cameras; 73 reflective markers distributed across joints
- Idealized linguistic gold standard baseline
- Green-screen environment utilizing mock-up car seat
- Animates high-fidelity avatars
- Captures detailed manual and non-manual signals

In-Car Multimodal Capture

- 1 x FLIR Blackfly S; 1 x Orbbec Femto Bolt; 1 x Lenovo Tablet; 1 x Prompting Mobil Phone with web-app ensures synchronization
- Deployed in Jeep, Nissan, Renault vehicles, fisheye lens maximize F.O.V
- Non-frontal, side-mounted sensor recording perspectives
- Dynamic interior shadows and solar noise
- Restricted cabin space limits signing scale
- Native signers in realistic geometry

Preliminary Data Analysis



- Multimodal fusion ensures stability during torso occlusions
- Infrared (IR) overcomes dynamic in-car lighting noise
- Depth maps enable effective hand-to-background segmentation
- Point clouds provide precise 3D trajectory hand tracking
- Synchronized multimodal streams provide a robust ground truth for training and this serves as first step toward robust in-the-wild SLR models

Potential Research Directions

- Camera placement optimization to maximize the visibility of facial and hand movements
- Physical baseline comparison (Lab vs Car)
- Evaluation of multimodal stability versus traditional monocular systems
- Simulation of non-frontal in-cabin angles using MoCap avatars
- Analysing the impacts of car seat and seatbelts on natural signing volume
- Testing model robustness against dynamic environmental noise

Conclusion

- Established first specialized in-car Libras resource and this provide a comprehensive research foundation
- Recorded 1.5M+ frames across multiple vehicles
- Systematically documented sign language production within constrained cabin environments
- Developed in direct collaboration with Deaf researchers
- This corpus enables future domain adaptation and HMI research
- Baseline for inclusive assistive transportation technologies
- Empirical resource for reducing communication barriers

Scan for full paper and technical specs.