Detecting and Classifying Hands in Social and Driving Contexts

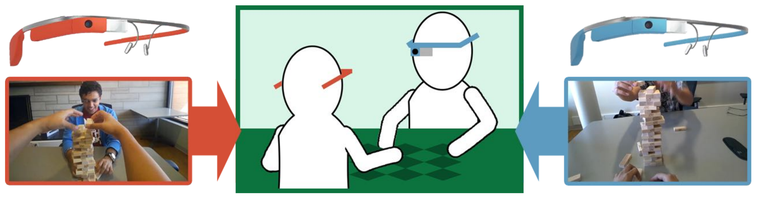
Sven Bambach, Stefan Lee, David Crandall, School of Informatics and Computing, Indiana University
Chen Yu, Department of Psychological and Brain Sciences, Indiana University

- We are interested in automatically **analyzing complex and dynamic interactions** from **first-person views**.
- To do this, we need to **robustly track hands** and **distinguish hand types** (my hands vs. your hands or left vs. right hands).
- We present an approach that **detects, distinguishes and segments hands** in real-world interactions using **visual features created through deep learning**.

- Hand detection and classification can provide helpful feedback to analyze **driver behavior** in active **safety systems for cars**.
- We demonstrate that our technique generalizes to such scenarios by applying it to the **VIVA 2015 challenge**.

1. Motivation

- We study egocentric hand detection, identification, and segmentation of interacting people in **realistic settings**.
- Evaluate the potential of **deep hand appearance models** to detect different hand poses and types.



2. Data Collection

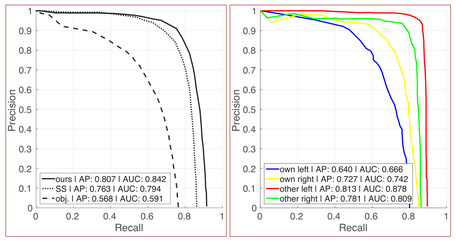
- Recorded synchronized first-person video from interacting subjects, using two **Google Glasses.**
- Four different actors, four activities, at three locations, for 4x4x3 = **48 unique videos.**
- Annotated 4,800 random frames with **pixel-level ground truth for 15,053 hands.**

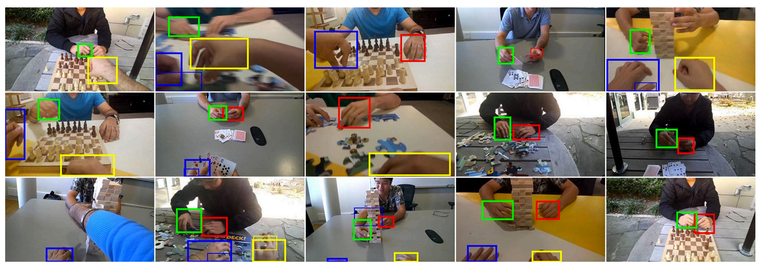
*Sample frames from our dataset. **Left**: Ground truth hand masks superimposed on sample frames, where colors indicate hand types. **Right:** Random subset of cropped hands according to ground truth segmentations.*

3. Hand Detection

- In first-person video, there are strong **spatial biases to hand location and size** for most natural actions.
- We use a **lightweight region proposal** technique that samples windows based on **skin color** and **spatial location**, yielding **better coverage** than other methods like “selective search” or “objectness.”
- We apply **convolutional neural networks (CNNs)** trained for a 5-way classification task between **own hands** (left/right), **other hands** (left/right), and background.

*Precision-recall for hand detection. **Left:** Results compared with other region-proposal methods. **Right:** Results for detecting four different hand types.*





Random detection results of own left, own right, other left and other right..

4. Segmenting Hands

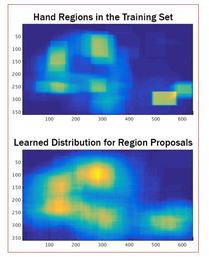
- We use our strong detections to initialize **GrabCut**, modified to use l**ocal color models** for hands and background.
- Yields **state-of-the-art results**.

***Top:** Segmentation examples on random frames. **Bottom:** Intersection/union results.*

Method	Own Hands		Other Hands		Average
	Left	Right	Left	Right	
Li et al.	0.395	0.478	0.534	0.505	0.478
Ours	**0.515**	**0.579**	**0.560**	**0.569**	**0.556**

VIVA Challenge

- We distinguish between the **driver’s left and right hands** and the **passenger’s left and right hands**.
- Cameras are either egocentric or mounted to the car (usually near the sunroof), imposing **spatial biases** on hand locations
- Our spatially sampled region proposals can cover **92% of hands** with only **500 windows** per frame.
- CNN evaluates proposed regions and distinguishes between hand types and background, yielding **state-of-the-art results** for **detection and classification**.



Some example detections for driver hands (left/right) and passenger hands (left/right). Most viewpoints observe people from the back/top, with some exceptions like first-person cameras or front-views from the A-pillar.

More Information:

- Full publications are currently under review (ICCV 2015, ICMI 2015)
- The “Egocentric Hand Interactions” dataset will be published online!

Challenge Results:

- A detailed quantitative evaluation can be found on the challenge website: **http://cvrr.ucsd.edu/vivachallenge/**

Acknowledgments: *This work was funded in part by the National Science Foundation (CAREER IIS-1253549), the National Institutes of Health (R01 HD074601 and R21 EY017843), the National Institute of Child Health and Human Development (5T32HD7475-17), Google, and the Indiana University Vice President for Research.*