

Spelling It Out: Real-time ASL Fingerspelling Recognition

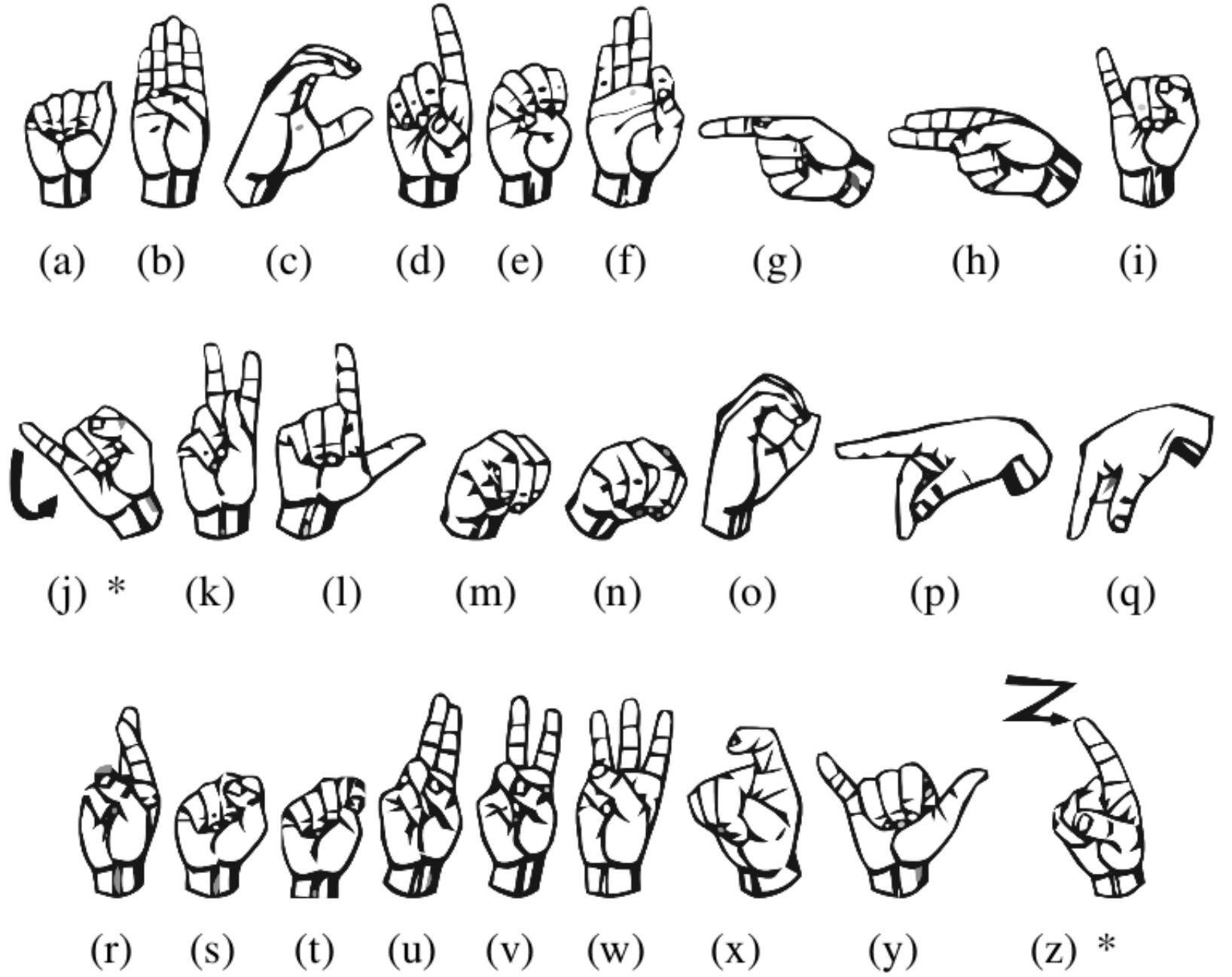


Nicolas Pugeault & Richard Bowden
{n.pugeault,r.bowden}@surrey.ac.uk
CVSSP, University of Surrey, UK



I. Introduction

This paper presents a method for ASL Fingerspelling recognition based on depth image (using Microsoft Kinect) and Machine Learning giving real-time prediction. This is embedded in an interactive fingerspelling user interface.

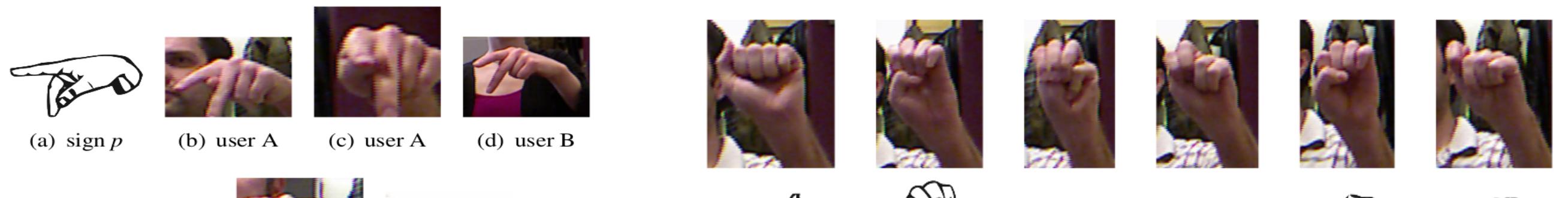


In American Sign Language (ASL), fingerspelling (signing all letters in a word) is done by 26 one-handed gestures, one per letter (see figure).

Note: we only recognize 24 out 26 signs because j and z involve motion rather than static hand shapes.

II. Dataset

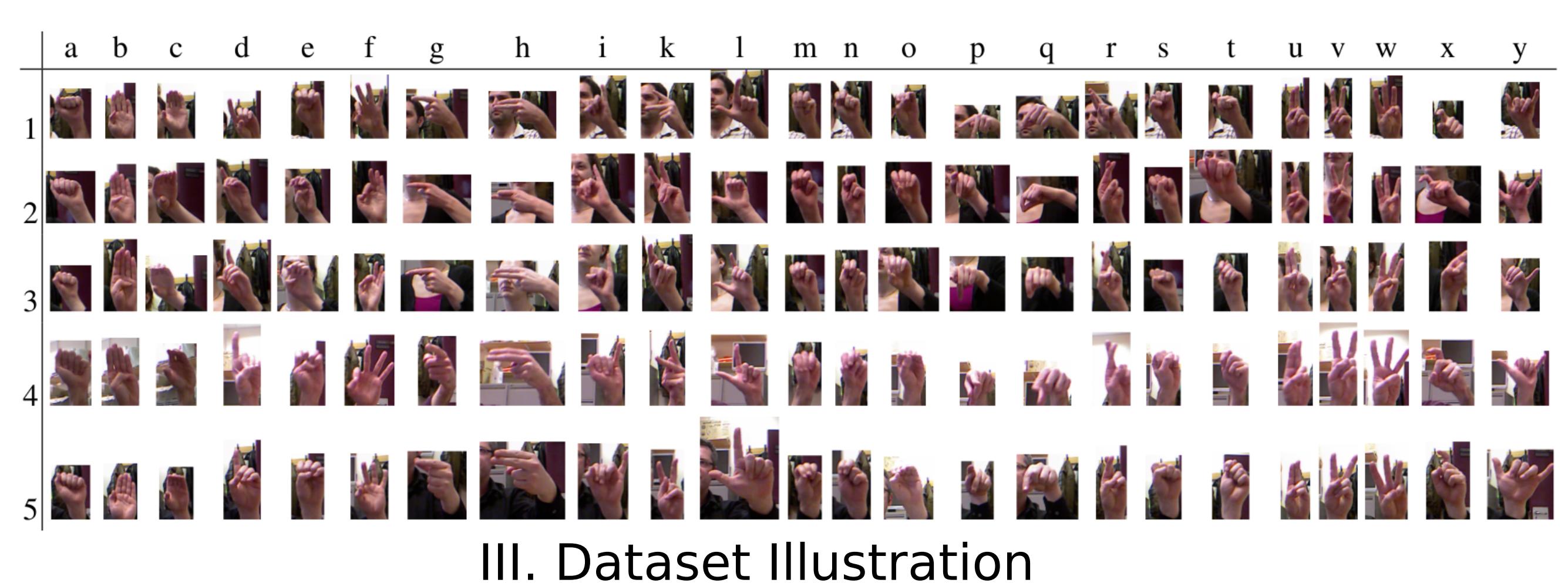
- 24 samples
- 4 different users * Hand segmentation:
-> Kinect/OpenNI/NITE
- 48,000 samples
- about 500 / sign



I. within class ambiguity

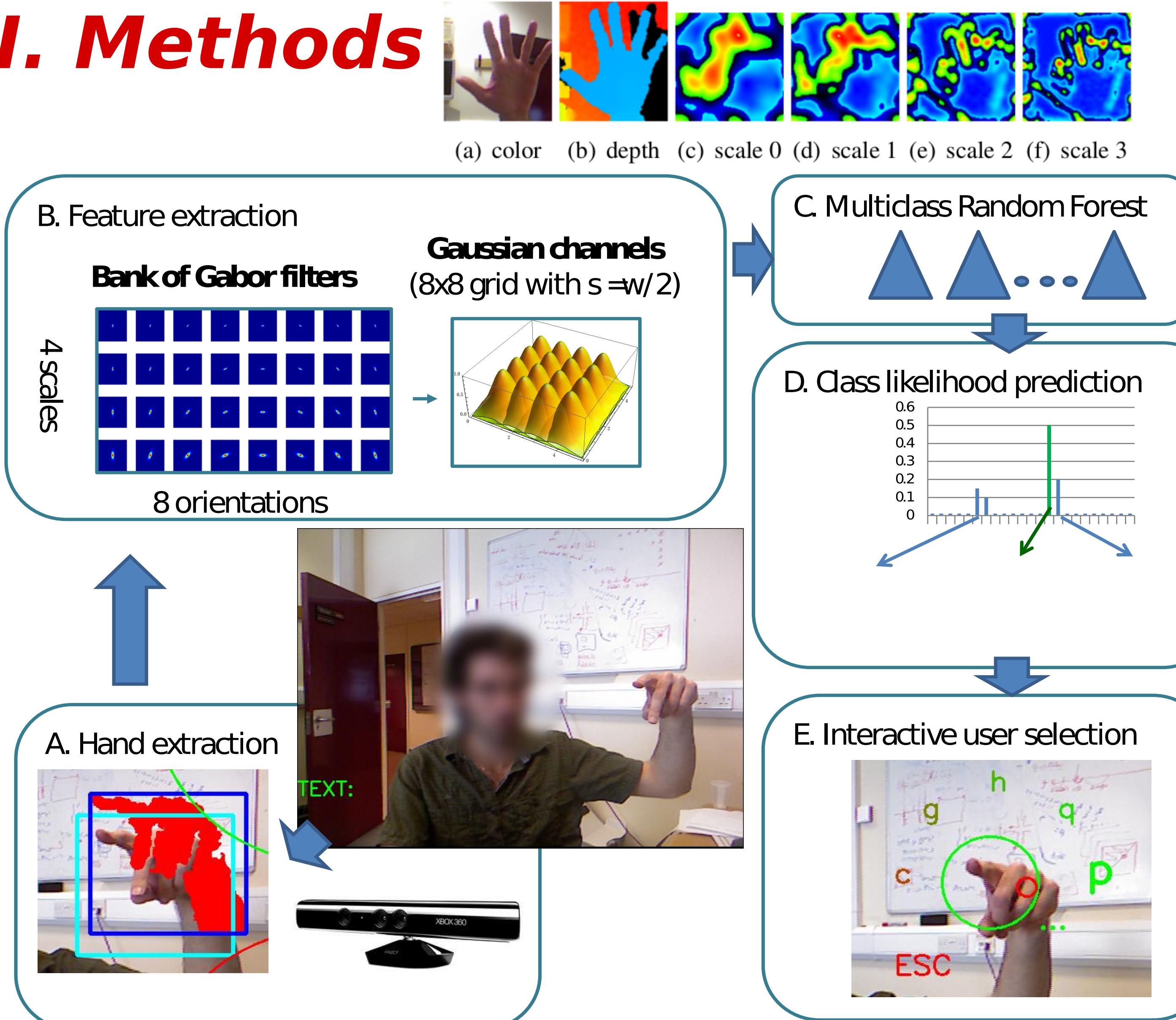


II. between class ambiguity



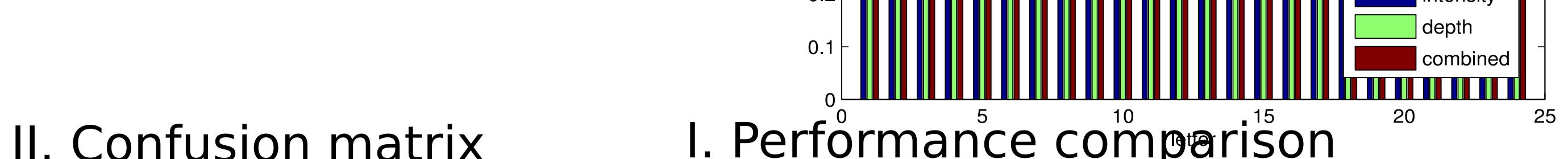
III. Dataset Illustration

III. Methods



IV. Results

- Half dataset for testing
- 100 multi-class trees
- depth 10
- about 12 FPS prediction



II. Confusion matrix

I. Performance comparison

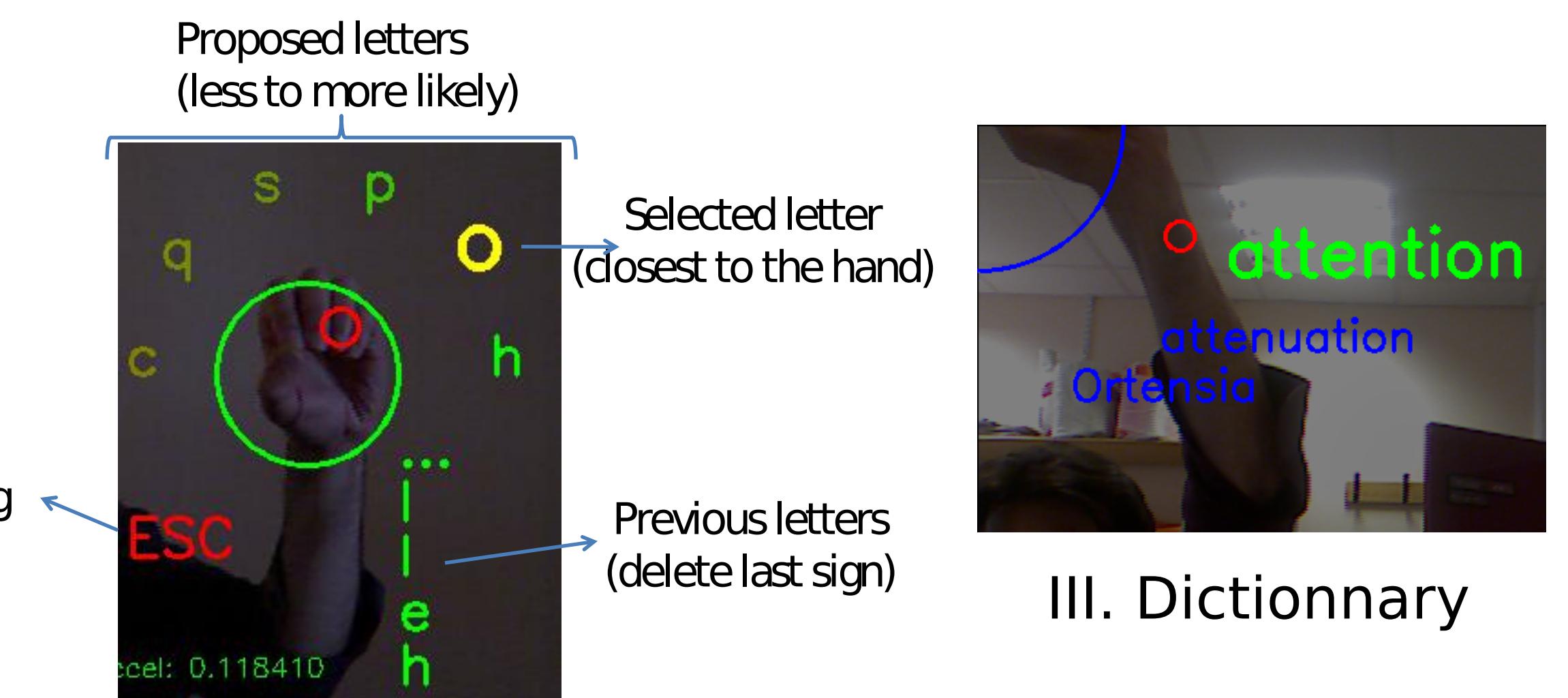
| | a | b | c | d | e | f | g | h | i | k | l | m | n | o | p | q | r | s | t | u | v | w | x | y |
|---|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|---|
| a | 0.75 | 0.05 | | | | | | | | | | | | | | | | | | | | | | |
| b | 0.03 | 0.83 | 0.03 | | | | | | | | | | | | | | | | | | | | | |
| c | | 0.57 | 0.13 | 0.03 | 0.03 | 0.03 | | | | | | | | | | | | | | | | | | |
| d | | | 0.37 | 0.13 | 0.03 | 0.03 | | | | | | | | | | | | | | | | | | |
| e | | | 0.07 | 0.03 | | 0.63 | | | | | | | | | | | | | | | | | | |
| f | | | | 0.30 | 0.10 | | 0.05 | 0.35 | | | | | | | | | | | | | | | | |
| g | | | | | 0.05 | 0.05 | 0.60 | | | | | | | | | | | | | | | | | |
| h | | | | | | 0.05 | 0.05 | 0.80 | | | | | | | | | | | | | | | | |
| i | | | | | | | 0.03 | 0.03 | 0.03 | | | | | | | | | | | | | | | |
| k | | | | | | | | 0.43 | 0.03 | | | | | | | | | | | | | | | |
| l | | | | | | | | | 0.87 | | | | | | | | | | | | | | | |
| m | | | | | | | | | | 0.13 | 0.10 | | | | | | | | | | | | | |
| n | | | | | | | | | | 0.17 | 0.10 | | | | | | | | | | | | | |
| o | | | | | | | | | | 0.30 | 0.13 | | | | | | | | | | | | | |
| p | | | | | | | | | | 0.07 | 0.10 | | | | | | | | | | | | | |
| q | | | | | | | | | | | 0.10 | 0.03 | | | | | | | | | | | | |
| r | | | | | | | | | | | 0.03 | 0.03 | 0.07 | | | | | | | | | | | |
| s | | | | | | | | | | | 0.13 | 0.03 | 0.03 | 0.10 | | | | | | | | | | |
| t | | | | | | | | | | | 0.33 | 0.13 | 0.03 | 0.07 | 0.13 | | | | | | | | | |
| u | | | | | | | | | | | | 0.17 | 0.03 | 0.03 | 0.03 | 0.03 | | | | | | | | |
| v | | | | | | | | | | | | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 | | | | | | | |
| w | | | | | | | | | | | | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 | | | | | | |
| x | | | | | | | | | | | | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 | | | | | |
| y | | | | | | | | | | | | 0.07 | 0.17 | 0.07 | 0.07 | 0.20 | 0.03 | 0.07 | 0.07 | 0.03 | 0.03 | 0.03 | 0.03 | |

V. Real-time user interface



I. Overall GUI and sensitive zones

Fingerspelling GUI



II. Fingerspelling & letter selection

VI. Summary & Conclusions

We have shown an efficient machine learning approach for hand-shape recognition, applied to American Sign Language fingerspelling recognition. This allowed a high-performance real-time prediction of signs, where the natural ambiguities between signs are addressed in a dynamic user interface.

References

- [1] Aspell. <http://aspell.net>.
- [2] OpenNI. <http://www.openni.org>.
- [3] Microsoft. Kinect. <http://www.xbox.com/kinect>
- [4] PrimeSense. NITE Middleware. <http://www.primesense.com/?p=515>
- [5] N. Pugeault and R. Bowden. Spelling it out: Real-time fingerspelling recognition (video). <http://www.youtube.com/watch?v=0tCGMhbTDmw>, 2011