

Appearance-Based Recognition of Signs in American Sign Language

Morteza Zahedi, Daniel Keyzers, and Hermann Ney

Lehrstuhl für Informatik VI, RWTH Aachen University, D-52056 Aachen, Germany

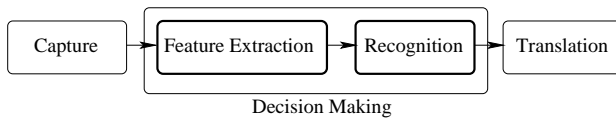
{zahedi,keyzers,ney}@informatik.rwth-aachen.de

Introduction

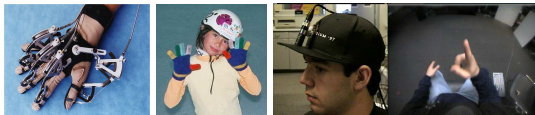
- Other recognition systems are based on special capture tools that are not easy to use in everyday life.
- In contrast, our system works with standard stationary cameras that are on fixed positions.
- For decision making, appearance-based features are extracted from the camera frame images.

General Architecture

- Architecture of the Sign Language translation system



- Capture tools
 - Data gloves
 - Colored gloves + camera
 - Stationary cameras (color and black/white cameras)
 - Wearable cameras



Example for capture tools (data glove, colored gloves and wearable camera)

- Translation with statistical models trained on parallel corpora.

Decision Making

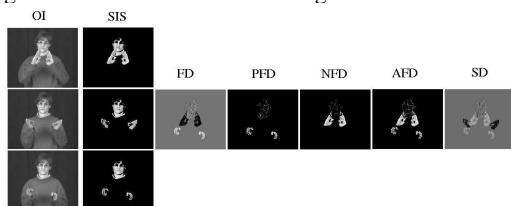


Appearance-Based Feature Extractor

HMM Classifier

Output

- Appearance-based features:
 - Original images (OI)
 - Skin color or skin intensity segmentation (SIS)
 - Downsampled images
 - First derivative (FD)
 - Positive, negative or absolute first derivative (PFD, NFD, AFD)
 - Second derivative (SD)
 - Using more than one camera and weights



- Hidden Markov model parameters:
 - Score function
 - Estimator function
 - Pooling
 - Topology of HMM
 - Distance function

Databases

- Boston ASL Database:
 - BOSTON10 (10 words with 110 utterances of ASL)
 - BOSTON50 (50 words with 483 utterances of ASL)
 - We considered 84 pronunciation for these words.
- Source: National center for Sign Language and gesture resources (Boston University)
- Using leaving one out method to train and classify
- Frame rate: 30 frames (312x242) per second (2 cameras)
- Signers: 1 man and 2 women

Camera 0:



Camera 1:



Experimental Results

BOSTON10 (using one camera):

Score function	ER(%)	
	Min. Seq. Length	Ave. Seq. Length
Gauss score (Standard deviation)	11	16
Laplace score (Mean deviation)	12	17
Laplace score (Mean deviation root)	29	38

BOSTON10 (using two cameras):

Camera	ER(%)
0	11
1	21
0,1	8
Weighted 0,1	7

BOSTON50 (using two cameras):

Method	ER(%)
HMM classifier	29.7
Consider Pronunciation	23.8
Using Tangent Distance	20.7

About 8% of our database are singleton utterances that occur only once in the corpus.

Conclusion and Future Work

Conclusion:

- Appearance-based features work well for Sign Language word recognition, and segmentation or tracking of the hands is not necessary.
- Use of tangent distance improves result of our HMM classifier.
- Using more than one camera improves our results.

Future work:

- Improving classification of single signs
 - Considering male and female signers
 - Using invariant features with respect to position and scale
 - Modelling of variability (tangent distance, image distortion model, ...)
- Continuous Sign Language recognition