



A Unifying Theory for Experimental Symbolomics

John Sample, Mike Test and Mary Try

Research Group for Experimental Symbolomics
CAS-MPG Partner Institute and Key Laboratory for Computational Biology
Shanghai Institutes for Biological Sciences, Shanghai, China

Overview

Sample poster with a more flexible/complex/interesting variant of the basic layout.

Introduction

- ▶ automatic sign language recognition system
- ▶ **necessary for communication** between deaf and hearing people
- ▶ **continuous** sign language recognition, **several** speakers, **vision-based** approach, **no special hardware**
- ▶ large vocabulary speech recognition (LVSR) system to obtain a textual representation of the signed sentences
- ▶ evaluation of speech recognition techniques on **publicly available sign language corpus**

Automatic Sign Language Recognition (ASLR)

- ▶ **similar to speech recognition**: temporal sequences of images
- ▶ important features
 - ▶ hand-shapes, facial expressions, lip-patterns
 - ▶ orientation and movement of the hands, arms or body
- ▶ HMMs are used to compensate time and amplitude variations of the signers
- ▶ **goal**: find the model which best expresses the observation sequence

Experimental Setup

Database

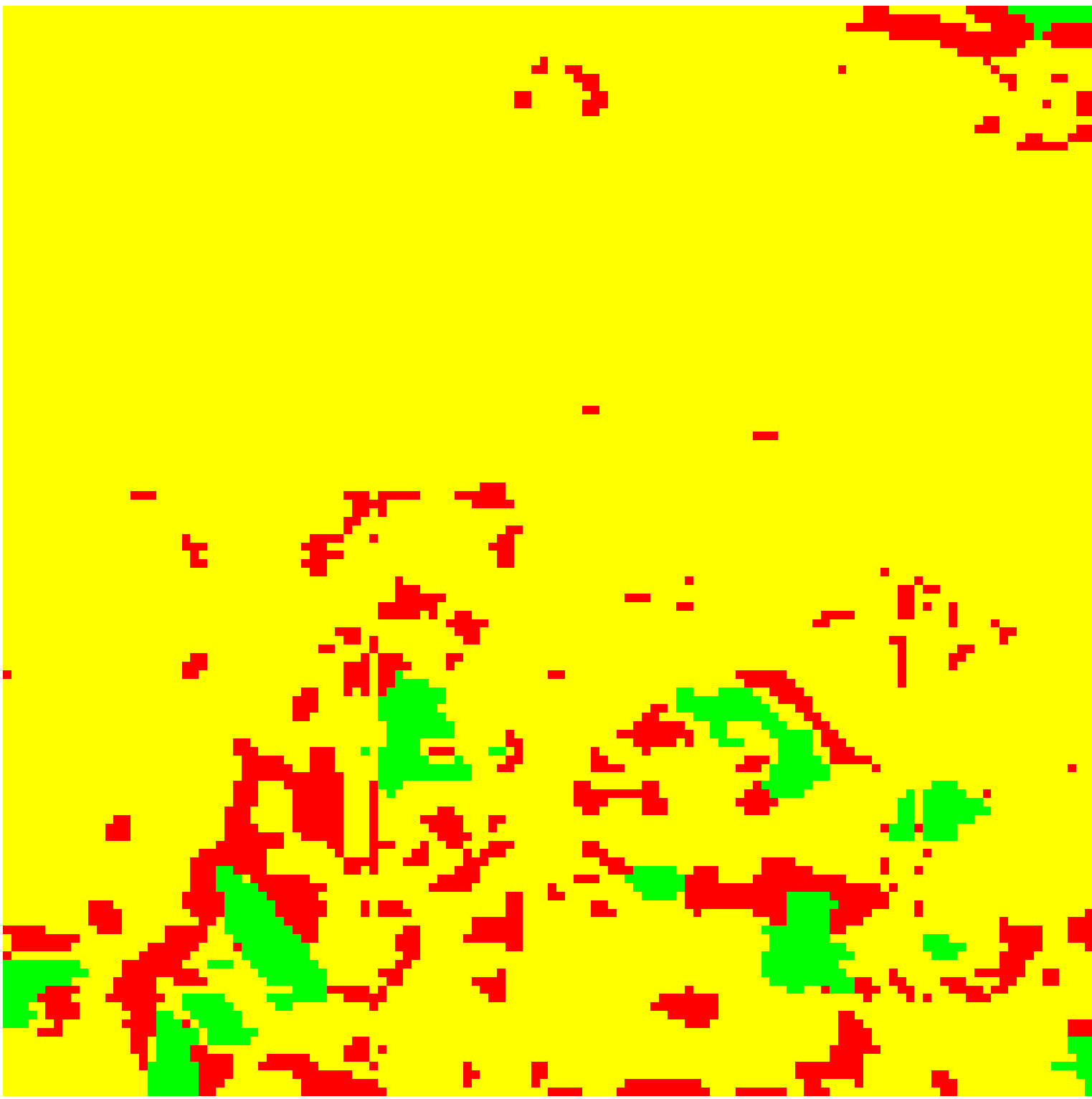
- ▶ system evaluation on the RWTH-BOSTON-104 database
 - ▶ **201 sentences** (161 training and 40 test sequences)
 - ▶ vocabulary size of **104 words**
 - ▶ 3 speakers (2 female, 1 male)
 - ▶ corpus is annotated in glosses

Problems

- ▶ 26% of the training data are **singletons**
- ▶ simple sentence structure
- ▶ one out-of-vocabulary (OOV) words with whole-word models

Differences in Comparison to ASR

- ▶ simultaneousness
- ▶ signing space
- ▶ environment
- ▶ speakers and dialects
- ▶ coarticulation and movement epenthesis
- ▶ silence
- ▶ whole-word models and sub-word units



Feature Selection and Model Combination

Feature Selection

- ▶ **concatenation** of appearance-based and manual features
- ▶ **sliding window** for context modeling
- ▶ **dimensionality reduction** by PCA and/or LDA

Model Combination

- ▶ **log-linear combination** of independently trained models
- ▶ profit from independent alignments (e.g. performing well for long and short words)
- ▶ profit from different feature extraction approaches

System Overview

Visual Modeling (VM)

- ▶ related to the acoustic model in ASR
- ▶ HMM based, with separate GMMs, globally pooled diag. covariance matrix
- ▶ monophone whole-word models
- ▶ pronunciation handling

Language Modeling (LM)

- ▶ according to ASR: LM should have a greater weight than the VM
- ▶ trigram LM using the SRILM toolkit, with modified Kneser-Ney discounting with interpolation