Hand in Hand: Automatic Sign Language to English Translation

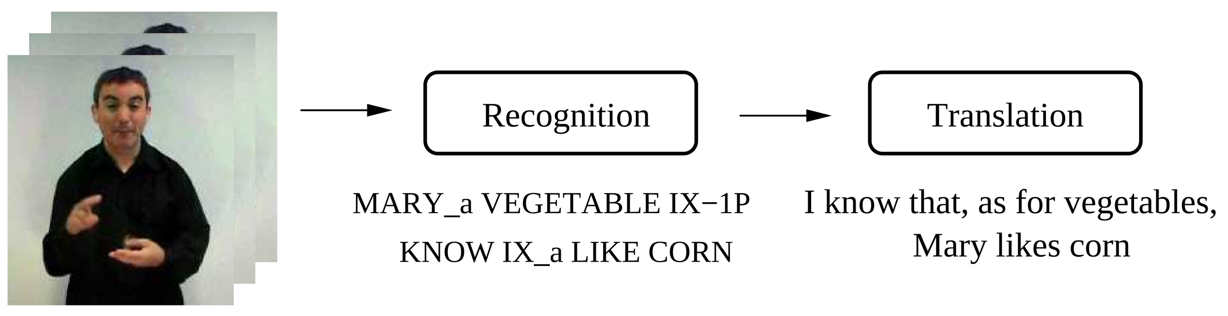
Daniel Stein, Philippe Dreuw, Hermann Ney
Human Language Technology and Pattern Recognition, RWTH Aachen University
Sara Morrissey, Andy Way
National Centre for Language Technology, Dublin City University

Introduction

- problem: all approaches in sign language research work on an intermediate language
- sign language in machine recognition:
  - input: video of signing person
  - output: semantic sign language representation (e.g. glosses)
- sign language in machine translation:
  - input: semantic sign language representation
  - output: written language (i.e. English)
- not directly intelligable by either hearing or deaf people
- incorporating statistical machine translation (SMT) on top of the recognition process:
  - converts glosses into written English
  - works even for very small corpora
  - data derived during the recognition can be used as additional knowledge source

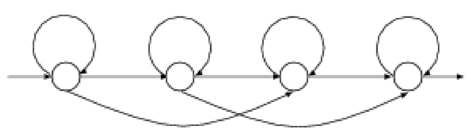
Intermediate Notation

- sign languages lack a formally adopted writing system
- syntactic representations describe handshape, location and movement of a sign
- glosses are a semantic representation of sign language
  - conventionally transcribed in the upper case stem form of the local spoken language
  - includes spatial and non-manual information

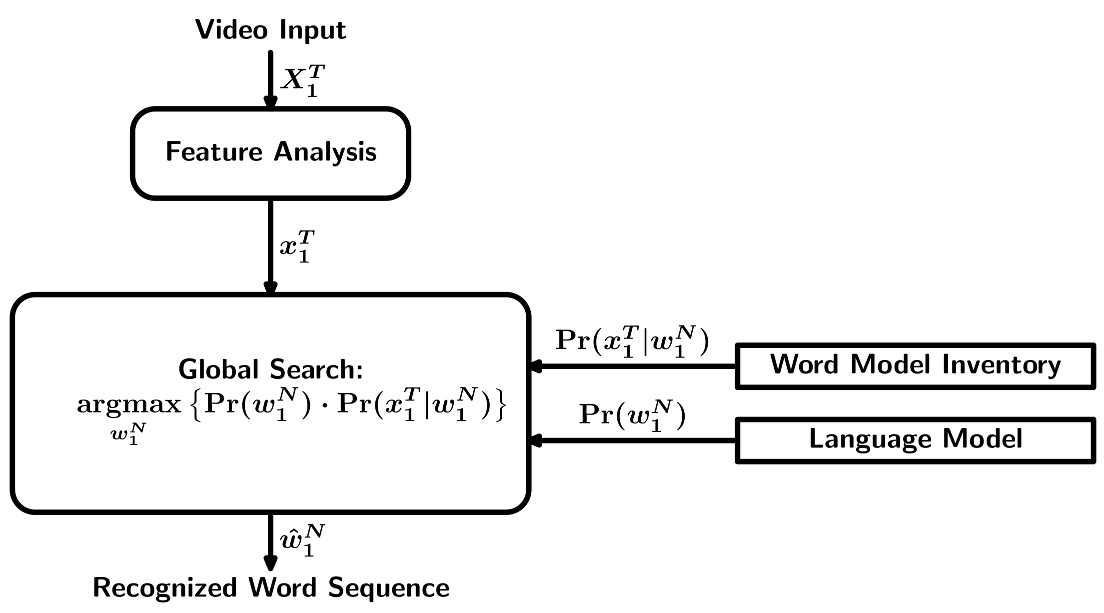


Sign Language Recognition

- a sign/gesture is a sequence 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
- to classify an observation sequence X_1^T, we use the Bayesian decision rule:



Tracking

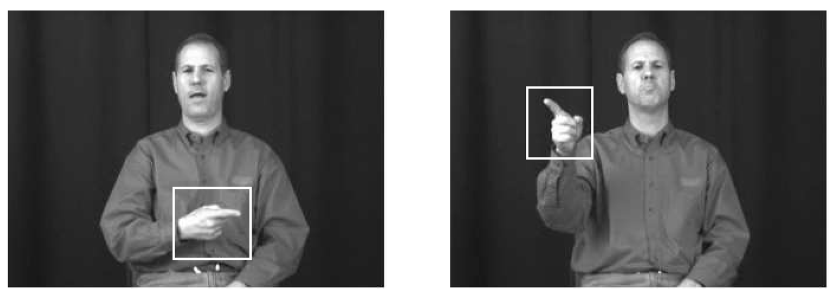
- tracking is done at the end of a sequence by tracking back the decisions to reconstruct the best path
- the best path is the path with the highest score wrt. a given scoring function

Sign Language Translation

- state-of-the-art phrase-based statistical machine translation system
  - for a recognized sequence f_1^J we maximize a translation probability for target sentences e_1^I
  - log-linear combination model:

$$p(e_1^I|f_1^J) = \frac{\exp\left(\sum_{m=1}^{M} \lambda_m h_m(e_1^I, f_1^J)\right)}{\sum_{\tilde{e}_1^I} \exp\left(\sum_{m=1}^{M} \lambda_m h_m(\tilde{e}_1^I, f_1^J)\right)}$$

  - set of different features h_m, scaling factors λ_m
  - trained with downhill simplex algorithm
- tracking positions of the sentences were clustered and their mean calculated
- for deictic signs, the nearest cluster according to the Euclidean distance was added as additional word information for the translation model

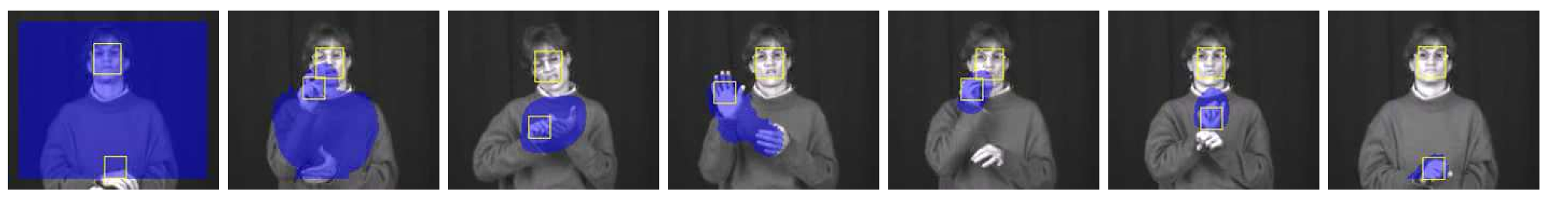


Experimental Results

- RWTH-Boston-104 database:
  - 161 training sentences, 40 test sentences

Experimental Results	WER[%]	PER[%]
recognition	17.9	-
translation	21.2	20.1
sign-to-speech	27.6	23.6

- RWTH-Boston-Hands database:
  - 1000 annotated frames, 2.3% tracking error rate
  - tracking of head and dominant-hand for ASLR



- enhancement with dominant-hand tracking features

Translation Features (different split)	WER[%]	PER[%]
without tracking	28.5	23.8
with tracking	26.5	23.5

Translation Example	
without tracking	John gives that man a coat
with tracking	John gives the man over there a coat.

Conclusion

- first data-driven automatic sign-language-to-speech translation system
- approach works for extremely small corpora typically encountered
- can be easily trained on new language pairs and new domains
- incorporation of the tracking data for the deictic words helps the translation system to discriminate between
  - distinctive article,
  - locative reference or
  - discourse entity reference

Outlook

- stemming of the glosses (i.e. leaving out the inflection)
- adding relevant features later in the translation
- model for all discourse entities
- handling spatial verb flexion, time information

Lehrstuhl für Informatik 6 – Computer Science Department – RWTH Aachen University, Germany
National Centre for Language Technology – School of Computing – Dublin City University, Ireland
{stein,dreuw,ney}@cs.rwth-aachen.de – {smorri,away} @computing.dcu.ie
http://www-i6.informatik.rwth-aachen.de--http://www.nclt.dcu.ie/mt/