Pattern Recognition 47 (2014) 1294–1308

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

# Towards subject independent continuous sign language recognition: A segment and merge approach

W.W. Kong [a], Surendra Ranganath [b,*]

[a] Department of Electrical & Computer Engineering, National University of Singapore, 4 Engineering Drive 3, Singapore 117576, Singapore
[b] Department of Information Science and Engineering, Sri Jayachamarajendra College of Engineering, Mysore 570002, India

ARTICLE INFO

*Article history:*
Received 2 June 2012
Received in revised form
12 September 2013
Accepted 21 September 2013
Available online 30 September 2013

*Keywords:*
Gesture recognition
Sign language recognition
Signer independence
Bayesian network
Conditional random field (CRF)
Support vector machine (SVM)
Semi-Markov CRF
Hidden Markov model (HMM)

ABSTRACT

This paper presents a segment-based probabilistic approach to robustly recognize continuous sign language sentences. The recognition strategy is based on a two-layer conditional random field (CRF) model, where the lower layer processes the component channels and provides outputs to the upper layer for sign recognition. The continuously signed sentences are first segmented, and the sub-segments are labeled *SIGN* or *ME* (movement epenthesis) by a Bayesian network (BN) which fuses the outputs of independent CRF and support vector machine (SVM) classifiers. The sub-segments labeled as *ME* are discarded and the remaining *SIGN* sub-segments are merged and recognized by the two-layer CRF classifier; for this we have proposed a new algorithm based on the semi-Markov CRF decoding scheme. With eight signers, we obtained a recall rate of 95.7% and a precision of 96.6% for unseen samples from seen signers, and a recall rate of 86.6% and a precision of 89.9% for unseen signers.



## 1. Introduction

In recent years, there has been increasing interest in developing automatic sign language recognition systems to facilitate communication between the deaf and hearing people. In manual signing, four components are used to compose signs, namely, handshape, movement, palm orientation and location; the systematic change of these components produces a large number of different sign appearances. The appearance and meaning of basic signs are well-defined in sign language dictionaries; however, in practice, variations arise due to regional, social, and ethnic factors, and also from gender, age, education and family background. This can lead to significant variations in manual signs performed by different signers, and pose challenging problems for developing robust computer-based sign language recognition systems.

### 1.1. Variations and movement epenthesis in manual signing

Variations which appear in the basic components, i.e. handshape, movement, palm orientation and location, are classified as phonological variation by linguists. Some handshapes are naturally close to each other; for example, the signs with handshapes “S” and “A” in American sign language (ASL) can look very similar if they are signed loosely. Also, some handshapes may be used interchangeably in certain signs, for example, signs such as FUNNY, NOSE, RED and CUTE are sometimes signed with or without thumb extension [1]. Studies in [2] show that ASL signs with handshape “1” (index finger extended, all other fingers and thumb closed) are very often signed as signs with handshape “L” (thumb and index extended, all other finger closed) or handshape “5” (all fingers open).

Locations of signs may also change. For example, the ASL sign for KNOW is prescribed to be signed at the forehead, but it is frequently signed at a lower position near the cheek. In [2], it was found that younger signers tend to make these signs below the forehead more often than older signers. Also, men tend to lower the sign location more than women. The movement path and palm orientation of a sign may also be modified; for example, signs with straight line movement can often be signed with arc-shaped movement or with palm orientation changing from palm-down to palm-left. Assimilation of handshape, movement, palm orientation and location also occur in compound signs. This refers to a process when two signs forming a compound sign begin to look similar. Some other phonological variations include deletion of one hand in a two-handed sign and hand contact.

There can be systematic variations present in the grammatical aspect of sign language. Typological variations concerning sign order

* Corresponding author. Tel: +91 900 8400 711.
*E-mail addresses:* wwn.kong@gmail.com (W.W. Kong), surendra@iitgn.ac.in (S. Ranganath).

0031-3203/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.patcog.2013.09.014

also occur where signs are arranged differently in sentences. Lastly, some signs can be made in unpredictable ways. For example, in [2] it was reported that the sign for PIZZA was fingerspelled by some, others signed it iconically as a person taking a bite out of a pizza, or as a round plate on which pizza is served. These variants of the sign do not share handshapes, locations, orientations and movements.

The above variations are related to linguistic aspects, and a sign language recognition system involving multiple signers must robustly handle these variations. In addition, physical variations (e.g. hand size, body size, length of the arm, etc. of the signers) also contribute to the complexity of building a robust recognition system.

Movement epenthesis (ME), the transition segment which connects successive signs, is formed when the hands move from the ending location of one sign to the starting location of the next sign, and does not carry any sign information. Linguistic studies of ME in the literature are limited and it does not have a well-defined lexicon. Perlmutter [3] also showed that ME had no phonological representation. As a connecting segment between signs, its starting and ending locations would depend on the preceding and succeeding signs, and its duration could even be comparable to that of a sign segment. Also, variations in adjacent signs may affect the ME, and it is possible that the variations in ME are comparable to variations in sign. As there are no well-defined rules for making such transitional movements, dealing with ME adds significant complexity to the task of recognizing continuous signs. This problem needs to be addressed explicitly for robust sign language recognition. It must be noted that ME is a different phenomenon from co-articulation in speech; co-articulation does occur in sign language, and manifests itself in some signs as hold deletion, metathesis and assimilation [2].

### 1.2. Motivation and scope

For sign language communication to be natural and effective, deviations from textbook definitions of sign can be expected. Hence, a practical sign language recognition system must be robust to these variations across signers. In the literature, most works which deal with signer independence issues consider hand postures or isolated signs [4] but works on signer independence with continuously signed sentences are limited. Some alternative approaches rely on an adaptation strategy, where a trained system is adapted to a new signer by collecting a small amount of signer specific data. Though adaptation is a reasonable approach, a truly signer independent system would be ideal.

Many recent works have considered recognition of continuously signed sentences, but their main focus has been on obtaining high recognition accuracy and scalability to large vocabulary. These are important problems to consider; however, several of these works report results based on only one signer. We consider recognizing continuously signed sentences from several signers, with resulting increase in inter-signer variations. In this paper we consider the phonological variations in sign language, i.e. variations in handshape, movement, palm orientation and location, arising from natural signing. We also include directional verbs which exhibit variation in grammatical aspect, and variations in sign order which can occur in natural signing.

In addition, inter-signer variations in ME also pose a challenge for accurate sign recognition. However, many works either neglect it or pay no special attention to the problem. In works that do consider it explicitly, the common approaches are either to model ME explicitly, or assume that the ME segments can be absorbed into adjacent signs. In this paper, we account for ME explicitly, though without elaborate modeling of these “extraneous” segments.

In the following, Section 2 summarizes related works. The overview of our proposed strategy for handling signer variation and ME is described in Section 3. Section 4 presents the feature representation used in the recognition framework. In Section 5, we discuss the strategy to deal with ME and present a conditional random field/support vector machine (CRF/SVM) and Bayesian network (BN) based classifier to discriminate between sign and ME. Sections 6 and 7 describe the complete recognition framework based on a two-layer CRF model and its decoding algorithm, respectively. Experimental results are presented in Section 8, and include comparisons with hidden Markov models (HMMs) along with results, analysis and discussion. Lastly, Section 9 gives the conclusions of this paper and suggestions for future work.

## 2. Related works

Recently, some works have considered the ME problem. Yang and Sarkar [5,6] adopted an enhanced level building algorithm (eLB) which was used to simultaneously segment and match signs to continuous sign language sentences. ME was automatically discarded during the matching process. They enhanced the classical level building algorithm [7] based on dynamic programming, and coupled it with a trigram grammar, to obtain 83.0% sign level recognition in [5]. Further experiments in [6] on ASL data sets showed that their approach outperformed CRFs and latent dynamic CRF-based approaches. In the works by Lee et al. [8–12], sign spotting in continuously signed sentences was used to recognize signs. They first trained a CRF model with only sign samples. Then, a threshold model with CRF was derived by adding the label for non-sign patterns by using the weights of the state and transition feature functions of the original CRF. ME segments were bypassed automatically. Testing on continuous ASL sentences consisting of 48 signs yielded 87.0% spotting rate and 93.5% recognition rate on the spotted isolated signs. Later extensions to spot signs and fingerspellings simultaneously using hierarchical CRFs [11,12] yielded 83.0% and 78.0% spotting rate for signs and fingerspellings, respectively. Kelly et al. [13,14] also proposed a parallel HMM threshold model to handle ME based on the threshold HMM proposed by Lee and Kim [15]. The key idea in threshold HMM is to use the likelihood as an adaptive threshold for selecting the proper gesture model. Kelly et al. [14] reported that 100 different types of ME and eight different signs were identified in experiments.

Works such as [5,6,10,14] used only signs for training and dealt with ME during the decoding process to avoid modeling the latter explicitly. However, Yang et al. [5,6] only used a single channel for processing and recognition, making the scheme vulnerable to signer variations, and limiting generalization to new signers. In their experiments with three signers, recognition results for a new signer were inconsistent. They reported accuracy of 80%, slightly more than 50% and less than 30% for three rounds of leave-one-out experiments with 10 sentences. The limited generalization could also be due to the generative modeling used for signs. Furthermore, their sign based modeling approach may not be scalable to large vocabulary compared to a phoneme-based approach. Both of the other works [10,14] were based on threshold models trained with only one signer, with parameters that were derived from the training data. However, finding good threshold values may be difficult when the problem is extended to several signers and the recognition framework may not perform robustly with new signers.

A practical continuous sign language recognition system needs to be signer independent; however, this has not received much attention in the literature. Presently, several continuous sign language recognition works, e.g. [10,16–21] and some of the works mentioned above mainly report results on a single signer. Two strategies for signer independent recognition are to (1) build a baseline recognition