Sequential Belief-Based Fusion of Manual and Non-Manual Signs

Oya Aran, Thomas Burger, Alice Caplier, Lale Akarun

▶ To cite this version:

HAL Id: hal-00327833
https://hal.science/hal-00327833

Sequential Belief-Based Fusion of Manual and Non-Manual Signs

Oya Aran[1], Thomas Burger[2], Alice Caplier[3], Lale Akarun[1]

[1]Dep. of Computer Engineering, Bogazici University 34342 Istanbul, Turkey
aranoya@boun.edu.tr, akarun@boun.edu.tr
[2]France Telecom R&D, 28 ch. Vieux Chêne, Meylan, France
thomas.burger@orange-ftgroup.com
[3]GIPSA-lab, 46 av. Felix Viallet, Grenoble, France
alice.caplier@lis.inpg.fr

Abstract. This work aims to recognize signs which have both manual and non-manual components by providing a sequential belief-based fusion mechanism. We propose a methodology based on belief functions for fusing extracted manual and non-manual information in a sequential two-step approach. The belief functions based on the likelihoods of the HMMs are used to decide whether there is an uncertainty in the decision of the first step and the uncertainty clusters. Then, we proceed to the second step, which utilizes only the non-manual features on the identified clusters, only if there is an uncertainty.

Keywords. Sign language recognition, manual and non-manual signs, hidden Markov models, belief functions

1   Introduction

Sign language is the natural communication medium of deaf people. Sign languages are visual languages and the message is contained not only in hand motion and shapes (manual signs, MS) but also in facial expressions, head/shoulder motion and body posture (non-manual signs, NMS). Most of the Sign Language Recognition (SLR) systems concentrate on hand gesture analysis only. However, without integrating NMS, it is not possible to extract the whole meaning of the sign. There are only a couple of studies that integrate MS and NMS for SLR (see [1] for a review). Current multimodal SLR systems either integrate lip motion and hand gestures, or only classify either the facial expression [2] or the head motion [3].

We propose a methodology for integrating manual and non-manual information in a sequential approach. The methodology is based on (1) identifying the level of uncertainty of a classification decision, (2) identifying sign clusters, and (3) identifying the correct sign based on MS and NMS. Section 2 explains our sequential belief-based fusion methodology and Section 3 gives the results of our experiments.

2 Sequential Belief Based Fusion

The sequential belief based fusion technique consists of two classification phases where the second classification phase is only applied when necessary (see Fig. 1). The necessity of applying the second phase is given by the belief functions defined on the likelihoods of the first bank of HMMs. The eventual uncertainty calculated from those beliefs is evaluated and resolved via the second bank of HMMs. In this setup, the assumption is that the HMMs of the first bank are general models which are capable of discriminating all the classes up to some degree. The HMMs of the second bank are specialized models and can only be used to discriminate between a subset of classes, among which there is an uncertainty. These uncertainties between classes are used to identify the sign clusters inside which the second bank of HMMs are capable of discriminating among individual signs.
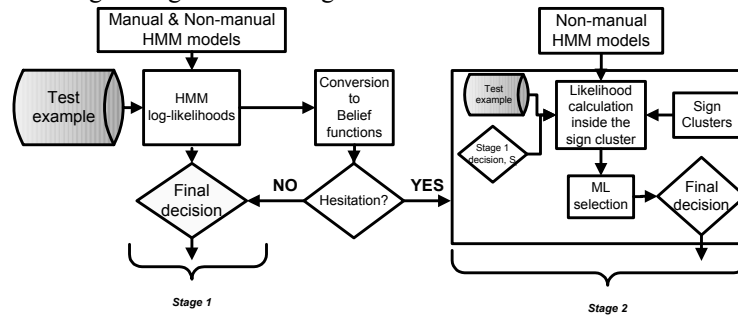

Fig. 1. Sequential belief-based fusion flowchart

2.1 Automatic Clustering for Sequential Fusion

We define a sign cluster as a group of signs which are similar and the differences are either based on NMS or variations of the MS. Our automatic cluster identification method is based on belief formalism. We propose to use the hesitation matrix for this purpose. The cluster identification is done by applying 7-fold cross-validation on the training data. The hesitation matrices of each fold are combined to create a joint matrix, which is used to identify the clusters.

3 Experiments

The experiments are conducted on eNTERFACE'06 sign language database [4]. The signs in the eNTERFACE'06 ASL Database are selected such that they include both manual and non-manual signs. There are eight base signs that represent words and a total of 19 variants which include the variations of the base signs in the form of NMS. Fig. 2 shows the signs CLEAN and VERY CLEAN. A single web camera with 640x480 resolution and 25 fps is used for the recordings. The database is collected

from eight subjects, each performing five repetitions of each sign where 532 examples are used for training and 228 examples for reporting the test results.

Fig. 2. Sign CLEAN and VERY CLEAN. The difference between these two signs is the existence of NMS, motion of the head.

Table 1. Classification performance

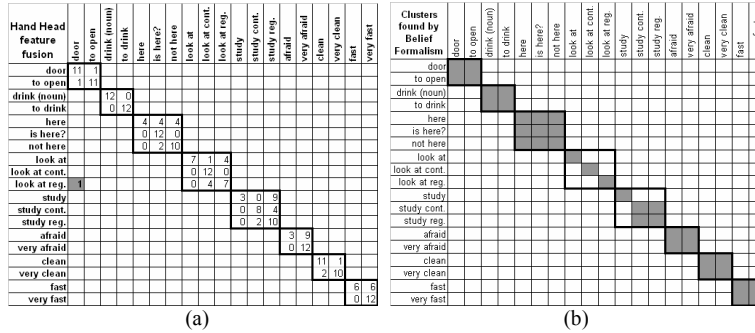| Models used	Fusion method	Cluster identification	Test Accuracy
HMM_M	No fusion	-	67.1 %
$HMM_{M\&N}$	Feature fusion	-	75.9 %
$HMM_{M\&N} \dashrightarrow HMM_N$	Sequential *belief-based* fusion	Automatic via uncertainties	**81.6 %**

Fig. 3. (a) Confusion matrix of $HMM_{M\&N}$: 99.5% base sign accuracy, 75.9% total accuracy. **(b)** Sign clusters identified by automatic belief-based clustering. Clusters are shown row-wise, where for each sign row, the shaded blocks show the signs in its cluster. Base sign and its variations are grouped and shown in bold squares.

Since we concentrate on the fusion step in this paper, we have directly used the processed data from [4]. Sign features are extracted for both MS and NMS. For hand motion analysis, the center of mass (CoM) of each hand is tracked and filtered by a Kalman Filter. The posterior states of each Kalman filter (position and velocity) form the hand motion features. Hand shape features include the parameters of an ellipse fitted to the binarized hand image and statistics from a rectangular mask placed on top of the binarized hand image. For head motion analysis, the system detects rigid head motions such as head rotations and head nods. The orientation and velocity of the head and the quantity of motion is used as head motion features. Further details on feature extraction can be found in [4].

To model the MS and NMS and perform classification, we trained three different HMMs. The first one is trained for comparison purposes and the last two are for the first and second steps of our fusion method: (1) HMM_M uses only *manual* features; (2) $HMM_{M\&N}$ uses both *manual* and *non-manual* features, and (3) HMM_N uses only *non-manual* features.

The accuracies of different fusion techniques are summarized in Table 1. Although the time dependency and synchronization of MS and NMS are not that high, feature fusion ($HMM_{M\&N}$) still improves the classification performance (13% improvement over HMM_M) by providing extra features of NMS. However, NMS is not effectively utilized by $HMM_{M\&N}$. Nevertheless, it is important that the classification errors are mainly between variants of a base sign and out of cluster errors are very few, with 99.5% base sign accuracy (Fig.3a). We further improve the accuracy by sequential-belief based fusion: up to 81.6%. The improvement is mainly based on the possibility of accepting the first stage classification decision thanks to belief formalism and the robustness of the belief-based cluster identification (Fig.3b).

4   Conclusions

We have proposed a technique for integrating manual and non-manual signs in a sign language recognition system. A dedicated fusion methodology is needed to accommodate the nature of MS and NMS in sign languages. This fusion technique makes use of the fact that the manual information is the primary component and the non-manual information is the secondary component of a sign. The sequential fusion method processes the signs accordingly. The key novelties of our fusion approach are two-fold. The first novelty is the two stage decision mechanism which ensures that if the decision at the first step is without hesitation, the decision is made immediately. The second novelty is the clustering mechanism: the sign clusters are identified automatically at the training phase and this makes the system flexible for adding new signs to the database by just providing new training data.

References

1. Ong, S.C.W. and Ranganath, S., "Automatic Sign Language Analysis: A survey and the Future beyond Lexical Meaning", IEEE Transactions on PAMI, vol.27, no.6, pp.873-891, June 2005.
2. Ming, K.W., Ranganath, S., "Representations for Facial Expressions," Proc. of Int. Conf. on Control Automation, Robotics and Vision, vol. 2, pp. 716-721, Dec. 2002.
3. Erdem U.M., S. Sclaroff, S., "Automatic Detection of Relevant Head Gestures in American Sign Language Communication," Proc. Int. Conf. on Pattern Recognition, vol. 1, pp. 460-463, 2002.
4. Aran, O., Ari, I., Benoit, F., Campr, A., Carrillo, A.H., Fanard, P., Akarun, L., Caplier, A., Rombaut, M. & Sankur, B., "Sign Language Tutoring Tool", eNTERFACE 2006, The Summer Workshop on Multimodal Interfaces, Croatia, 2006