

Author's Accepted Manuscript

A belief-based sequential fusion approach for fusing manual and non-manual signs

Oya Aran, Thomas Burger, Alice Caplier, Lale Akarun

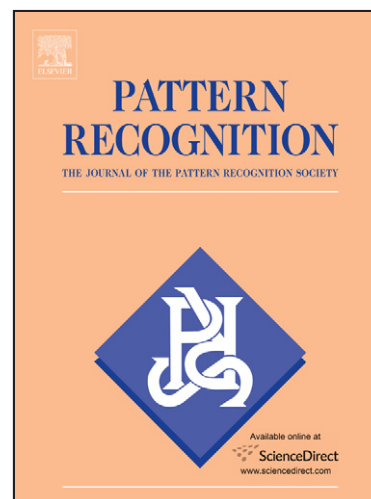
PII: S0031-3203(08)00401-9
DOI: doi:10.1016/j.patcog.2008.09.010
Reference: PR 3335

To appear in: *Pattern Recognition*

Received date: 3 August 2007
Revised date: 4 June 2008
Accepted date: 21 September 2008

Cite this article as: Oya Aran, Thomas Burger, Alice Caplier and Lale Akarun, A belief-based sequential fusion approach for fusing manual and non-manual signs, *Pattern Recognition* (2008), doi:[10.1016/j.patcog.2008.09.010](https://doi.org/10.1016/j.patcog.2008.09.010)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting galley proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



www.elsevier.com/locate/pr

A Belief-Based Sequential Fusion Approach for Fusing Manual and Non-Manual Signs

Oya Aran ^{a,*}, Thomas Burger ^b, Alice Caplier ^c, Lale Akarun ^a

^a*Dep. of Computer Engineering, Bogazici University 34342 Istanbul, Turkey*

^b*France Telecom R&D, 28 ch. Vieux Chêne, 38240, Meylan, France*

^c*GIPSA-lab, 46 avenue Félix Viallet, 38031 Grenoble cedex 1, France*

Abstract

Most of the research on sign language recognition concentrates on recognizing only manual signs (hand gestures and shapes), discarding a very important component: the non-manual signs (facial expressions and head/shoulder motion). We address the recognition of signs with both manual and non-manual components using a sequential belief-based fusion technique. The manual components, which carry information of primary importance, are utilized in the first stage. The second stage, which makes use of non-manual components, is only employed if there is hesitation in the decision of the first stage. We employ belief formalism both to model the hesitation and to determine the sign clusters within which the discrimination takes place in the second stage. We have implemented this technique in a sign tutor application. Our results on the eNTERFACE'06 ASL database show an improvement over the baseline system which uses parallel or feature fusion of manual and non-manual features: We achieve an accuracy of 81.6%.

Key words: Sign language recognition, manual and non-manual signs, Hidden Markov Models, data fusion, belief functions

1 Introduction

Sign languages are the natural communication media of hearing-impaired people all over the world. Like the spoken languages, they emerge spontaneously, and evolve naturally within deaf communities. Wherever deaf communities exist, sign languages develop, without necessarily having a connection with

* Corresponding Author, Tel:+90 212 3597183 Fax:+90 212 2872461
Email address: aranoya@boun.edu.tr (Oya Aran).

the spoken language of the region. American Sign Language, British Sign Language, Turkish Sign Language, and French Sign Language are different sign languages used by corresponding communities of hearing-impaired people. Sign languages are visual languages: the phonology makes use of the hand shape, place of articulation, and movement; the morphology uses directionality, aspect and numeral incorporation, and syntax uses spatial localization and agreement as well as facial expressions. The whole message is contained not only in hand motion and shapes (manual signs) but also in facial expressions, head/shoulder motion and body posture (non-manual signs). As a consequence, the language is intrinsically multimodal [1,2].

The problem of sign language recognition (SLR) can be defined as the analysis of all components that form the language and the comprehension of a single sign or a whole sequence of sign language communication. SLR is a very complex task: a task that uses hand shape recognition, gesture recognition, face and body parts detection, and facial expression recognition as basic building blocks. Hand gesture analysis [3] is very important for SLR since the manual signals are the basic components that form the signs. However, without integrating non-manual signs, it is not possible to extract the whole meaning of the sign. In almost all of the sign languages, the meaning of a sign can be changed drastically by the facial expression or the body posture accompanying a hand gesture. Current multimodal SLR systems either integrate lip motion and hand gestures, or only classify either the facial expression or the head movement. There are only a couple of studies that integrate non-manual and manual cues for SLR [4].

Initial studies on sign language recognition have concentrated on static signs, attempting to recognize either the finger spelling alphabet or some selected static signs. These studies were based on hand shape recognition, discarding the temporal information. Recognizing static signs is a complex problem as a result of the high degree of freedom of the human hand and it is an ongoing research topic [5]. In later studies, with increased processing capabilities of computers and camera speeds, researchers started to work on temporal data to recognize dynamic signs that include the motion of the hand. Among several methods for modeling temporal data, HMMs are used the most extensively and have proven successful in several kinds of SLR systems. Different HMM architectures are compared [6] and more specialized HMM architectures are proposed for modeling two handed signs [7].

Non-manual signs in sign language have only recently drawn attention for recognition purposes. Most of those studies attempt to recognize non-manual information independently, discarding the manual information. Some works only use facial expressions [8], and some use only the head motion [9], without considering the manual component. There are also studies that try to improve SLR performance by lip reading [10]. However, lip reading is not a component

of the sign language and can be omitted without any effect on the meaning. Contrarily, the analysis of non-manual signs is a must for building a complete SLR system: two signs with the same manual component and different non-manual components can have completely different meanings.

The dependency and correlation of manual and non-manual information during a recognition task must be further investigated. For the case of isolated signs, the manual and non-manual information coincide but the internal correlation and dependency are fairly low. For each isolated sign, there may or may not be a non-manual component. However, for continuous signing, the manual and non-manual components do not have to coincide synchronously and non-manual signs may cover more than one manual sign. This paper is focused on isolated signs, assuming that subjects perform a manual sign with or without an accompanying non-manual sign. Continuous signing is out of the scope of this paper.

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, i.e., groups of signs that include different non-manual components or variations of a base manual component, and (3) identifying the correct sign based on manual and non-manual information. Sections 2 and 3 give background information on Belief Functions, and Hidden Markov Models, respectively. Section 4 explains the sequential belief-based fusion technique and our methodology to assign belief values to our decisions and to calculate the uncertainty. In Section 5, we compare our fusion technique with other state of the art fusion techniques and give the results of experiments with detailed discussions.

2 Belief Functions and the Transferable Belief Model

Belief function formalism may be explained in many ways. Although not the most rigorous, the most intuitive way is to consider it as a generalization of probability theory which provides a way to represent hesitation and ignorance indifferently. This formalism (called the “evidential” formalism) is especially useful when the collected data is noisy or semi-reliable. In this section, we briefly present the necessary background on belief functions. Interested readers may refer to [11–14] for more information on belief theories.

Frame: Let Ω be the set of N exclusive hypotheses: $\Omega = \{\Omega_1, \Omega_2, \dots, \Omega_N\}$. Ω is called the frame. It is the evidential counterpart of the probabilistic universe.

Powerset: Let 2^Ω , called the powerset of Ω , be the set of all the subsets A of Ω , including the empty set: $2^\Omega = \{A/A \subseteq \Omega\}$

Belief function (BF): A belief function is a set of scores defined on 2^Ω and adds up to 1, in exactly the same manner as a probability function (PF) defined on Ω . Let $m(\cdot)$ be such a belief function. It represents our belief in the propositions that correspond to the elements of 2^Ω :

$$m : 2^\Omega \rightarrow [0, 1]$$

$$A \mapsto m(A) \text{ with } \sum_{A \subseteq \Omega} m(A) = 1$$

A focal element is an element of the powerset to which a non-zero belief is assigned. Note that belief can be assigned to non-singleton propositions, which allows modeling the hesitation due to the absence of knowledge between elements.

Dempster's rule of combination: To combine several belief functions into a global belief function, one uses the Dempster's rule of combination. For N BFs, $m_1 \dots m_N$, defined on the same hypothesis set Ω , the Dempster's rule of combination is defined as:

$$\bigcirc : \mathfrak{B}^\Omega \times \mathfrak{B}^\Omega \times \dots \times \mathfrak{B}^\Omega \rightarrow \mathfrak{B}^\Omega$$

$$m_1 \bigcirc m_2 \bigcirc \dots \bigcirc m_N \mapsto m_\bigcirc$$

with \mathfrak{B}^Ω , the set of BFs defined on Ω , and m_\bigcirc , the global combined BF, which is calculated as:

$$m_\bigcirc(A) = \sum_{A=A_1 \cap \dots \cap A_N} \left(\prod_{n=1}^N m_n(A_n) \right) \quad \forall A \in 2^\Omega \quad (1)$$

Decision making: After fusing several BFs, the knowledge is modeled via a single function over 2^Ω . There are alternative ways to make a decision from the knowledge of a BF [15–17]. A very popular method is to use the *Pignistic Transform* (PT), such as defined in the Transfer Belief Model [13]. One defines $\text{BetP}(\cdot)$ [13], the result of the PT of a BF $m(\cdot)$, as:

$$\text{BetP}(h) = \frac{1}{1 - m(\emptyset)} \sum_{h \in A, A \subseteq \Omega} \frac{m(A)}{|A|} \quad \forall h \in \Omega \quad (2)$$

where $|A|$ denotes the cardinality of A . The division by $1 - m(\emptyset)$ is a normalizing factor. We will use the following notation of conditioning to express this normalization: $\cdot | \Omega$. In BF formalism, it is possible to make decisions based on other assumptions. For instance, it is possible to associate a plausibility

to any element of the powerset, and then select the most plausible element [18,19]. The plausibility of an element is the sum of all the belief associated with the hypothesis which fails to refute it. Hence, the bigger the cardinality of an element of the powerset is, the higher its plausibility is. Then, deciding according to the plausibility measure is likely to lead to a decision on a set of hypotheses of high cardinality, including the entire Ω , which is finally an absence of decision. It may be wiser to prevent any decision making than making an inaccurate decision, especially in case of a decision with a huge cost of erring (juridic decision, for instance). Basically, these two stances, to make a bet, or to wait for a cautious decision are typically opposed in decision making. For sign language recognition, we have the necessity to make a decision on which a reasonable mistake is acceptable, but one needs to reject a bet when excessive amount of information is missing: when classifying a sign with both manual and non-manual information, we accept a part of indecision on the non-manual gesture, but we have to make a complete decision on its manual part (see next sections). Hence, we need to use an intermediate way to make a decision: We need to precisely set the balance between cautiousness and the risk of a bet.

We propose to define a new method based on the PT. We generalize the PT so that we can decide whether any focal element has a too large cardinality with respect to the amount of uncertainty we allow, or, on the contrary, it is small enough (even if it is not a singleton focal element), to be considered. We call this transformation as Partial Pignistic Transform (PPT).

Partial Pignistic Transform: Let γ be an uncertainty threshold, and let $2^{|\Omega|_\gamma}$ be the set of all elements of the frame with cardinality smaller or equal to γ . We call $2^{|\Omega|_\gamma}$ the γ^{th} frame of decision.

$$2^{|\Omega|_\gamma} = \{A \in 2^\Omega / |A| \in [0, \dots, \gamma]\} \quad (3)$$

Let $M_\gamma(\cdot)$ be the result by γ -PPT of a BF $m(\cdot)$. It is defined on $2^{|\Omega|_\gamma}$ as:

$$M_\gamma(\emptyset) = 0$$

$$M_\gamma(A) = m(A) + \sum_{B \supseteq A, B \notin 2^{|\Omega|_\gamma}} m(B) \frac{|A|}{\sum_{k=1}^{\gamma} \left[\binom{|B|}{k} \cdot k \right]}, \quad \forall A \in 2^{|\Omega|_\gamma} \setminus \emptyset \quad (4)$$

where B are supersets of A, and $|A|$ denotes the cardinality of A. Then, the decision is made by simply choosing the most believable element of the γ^{th} frame of decision: $D^* = \text{argmax}_{2^{|\Omega|_\gamma}} (M_\gamma)$. Note that, the 1-PPT is equivalent to the PT. As an illustration of all these concepts, let us consider the numerical example in Table 1. Assume that we want to automatically classify a hand gesture. The gesture can be the trace of one of the following shapes: square (S), circle (C) or triangle (T). The gesture is analyzed by two different sensors, each

Table 1
Numerical example for belief function usage

	\emptyset	S	C	T	$\{S, C\}$	$\{S, T\}$	$\{T, C\}$	$\{S, C, T\}$
m_1	0	0.5	0	0	0.5	0	0	0
m_2	0	0	0	0	0	0	0.4	0.6
$m_{\odot} = m_1 \odot m_2$	0.2	0.3	0.2	0	0.3	0	0	0
M_1	0.2	0.45	0.35	0	0	0	0	0
$M_1 \Omega$	0	0.56	0.44	0	0	0	0	0
M_2	0.2	0.3	0.2	0	0.3	0	0	0

giving an estimation of its shape. The observations of the sensors are expressed as beliefs, $m_1(\cdot)$ and $m_2(\cdot)$, and the powerset of hypotheses for the shape of the gesture is defined as $2^{\Omega} = \emptyset, S, C, T, \{S, C\}, \{S, T\}, \{T, C\}, \{S, C, T\}$. The two beliefs (m_1, m_2) are fused into a new belief (m_{\odot}) via the Dempster's rule of combination (Eq. 1). As the gesture has a single shape, the belief in union of shapes is meaningless from a decision making point of view: One needs to make a complete decision without hesitation (M_1), and the adapted frame of decision is the 1st frame of decision (Eq. 3) with 1-PPT (Eq. 4). In addition, if the ground truth is assumed to be represented in the frame, then the belief in the empty set is not meaningful and conditioning on Ω leads to the same result as the classical Pignistic Transform ($M_1|\Omega$). If the sensors are not precise enough to differentiate between two particular gestures, then, 2-PPT may be used (M_2). In this example, $M_2 = m_{\odot}$ as there is no uncertainty in the elements with cardinality three to share: $m_{\odot}(\{S, C, T\}) = 0$.

3 Hidden Markov Models

HMMs are among the most popular and powerful algorithms for sequence modeling and classification problems [20]. Among different kinds of HMM architectures, left-to-right HMMs with either discrete or continuous observations are preferred for their simplicity and suitability to the hand gesture and sign language recognition.

The elements of an HMM are prior probabilities of states, π_i , transition probabilities, a_{ij} and observation probabilities, $b_i(O_t)$ where $1 \leq i \leq N$ and N is the number of states. In the discrete case, observation probabilities are stored in a $N \times M$ matrix, M being the number of symbols in the alphabet. In the continuous case, observation probabilities are modeled by mixture of multivariate Gaussians.

$$b_i(O_t) = \sum_{k=1}^K w_{ik} \mathcal{N}(O_t; \mu_{ik}, \Sigma_{ik}) \quad (5)$$

where μ_{ik}, Σ_{ik} are component means and covariances, respectively, and K denotes the number of components in the mixture. V is the dimensionality of the observation vectors and O_t is the observation at time t .

For a sequence classification problem, one is interested in evaluating the probability of any given observation sequence, $O_1 O_2 \dots O_T$, given a HMM model, Θ . This probability, or the likelihood, $P(O|\Theta)$, of an HMM can be calculated in terms of the forward variable.

$$P(O|\Theta) = \sum_{i=1}^N \alpha_T(i) \quad (6)$$

where the forward variable, $\alpha_T(i)$, is the probability of observing the partial sequence $O_1 \dots O_T$ until the end of the sequence, T , and being in state i at time T , given the model Θ . The forward variable can be recursively calculated by going forward in time:

$$\alpha_1(j) = \pi_j b_j(O_1) \quad (7)$$

$$\alpha_t(j) = b_j(O_t) \sum_{i=1}^N \alpha_{t-1}(i) a_{ij} \quad (8)$$

For long sequences, the computation of the forward variable will exceed the precision range of the machine. Thus, a scaling procedure is needed to prevent underflow. The scaling coefficient, c_t is calculated as follows:

$$c_t = \frac{1}{\sum_{i=1}^N \alpha_t(i)} \quad (9)$$

The computation of $P(O|\Theta)$ must be handled differently since $\alpha_T(i)$ s are already scaled. $P(O|\Theta)$ can be calculated via the scaling coefficients. However we can only calculate the logarithm of P since P itself will be out of the precision range [20]:

$$\log(P(O|\Theta)) = -\sum_{t=1}^T \log c_t \quad (10)$$

4 Sequential Belief Based Fusion

In a SLR problem, where each sign is modeled by a generative model, such as an HMM, the classification can be done via the maximum likelihood approach, where the sign class of the HMM that gives the maximum likelihood is selected. However, this approach does not consider situations where the likelihoods of two or more HMMs are very close to each other. The decisions made in these kinds of cases are error-prone and further analysis must be made.

In HMM based SLR, each HMM typically models a different hypothesis for the sign to be recognized. Our purpose is to associate a belief function with these likelihoods. Then, it is possible to model these error-prone cases by associating high belief into the union of hypotheses. By analyzing the proportion of belief which is associated with the union of hypotheses, it is possible to decide whether the classification decision is certain or error-prone. Then, we propose the following process: If the analysis indicates significant uncertainty in the decision of the first classification step, a second classification step must be applied. This second classification step is only applied to classes among which the uncertainty is detected. The preliminary results of this methodology is presented in [21].

In the following sections, we first explain how to convert the HMM likelihoods to beliefs and how to introduce uncertainty using the calculated belief values. Then we explain how to use this uncertainty for sequential fusion.

4.1 Belief function definition from HMM log-likelihoods

We present a method to derive a belief function over the power set of classes from the HMM log-likelihoods calculated for each class. The purpose is to convert the information contained in the log-likelihoods into the evidential formalism, so that corresponding methods in data fusion and decision making are usable.

The general idea is to automatically copy the fashion in which an expert would associate a BF to a set of likelihoods: In the case of an HMM H^* having a significantly larger likelihood than the other HMMs, an expert is able to place her/his belief into H^* . On the contrary, s/he is able to share the belief among a set of HMMs which roughly have the same likelihoods.

A way to understand this is to see the human being as capable of considering simultaneously several pairwise comparisons and to apprehend their global interactions. We propose to analytically copy this behavior: Let us consider all possible pairs of HMMs involved, and associate an Elementary Belief Function

(EBF) to each of them. We assume that the higher the likelihood is, the more believable the corresponding sign. Then, the margin (the algebraic difference) between the likelihoods of two HMMs is a basis for local decision. This is inspired from our previous work [22], in which we proposed a scheme that uses Belief Theories with SVMs to improve the 1vs1 voting scheme for multi-class classification.

The next crucial step is to decide how to numerically associate a belief function over the power set of every couple. Under which values will margins be considered as hesitation-prone? We define each EBF over the powerset of the two HMMs involved. Then, the belief of each EBF is distributed over one HMM, the other HMM, and the hesitation among the two HMMs. We modify this partition so that the HMM which has the smaller value among the two has a zero-valued belief. So, we simply define the repartition between one HMM and the union of the two involved. This is what we call the hesitation distribution, b , and it is modeled on the behavior of an expert human being. Then, for each pair of HMMs (HMM_i , HMM_j) an EBF m_{ij} is defined as

$$\begin{aligned} m_{ij}(\{i, j\}) &= b, \quad b \in [0, 1] \\ m_{ij}(\{i\}) &= 1 - b \\ m_{ij}(\{j\}) &= 0 \end{aligned} \tag{11}$$

assuming that HMM_i gives a higher score than HMM_j . Then, the only point is to define the hesitation distribution, b , for each EBF.

We should keep in mind that, although the HMM scores are derived from probabilities, they are indeed log-likelihoods and it is not possible to compute the inverse and to go back to likelihoods because of the scaling operation carried out to prevent underflow of the probabilities when the sequences are long. Hence, if scaling is used, only the log-likelihood is defined, and not the likelihood itself, due to the limits of machine representation of numbers. [20].

This “conversion” problem is far more complicated when converting a real probability function into a belief function. The simple solution would be to convert the scores to probabilities by scaling and normalization operations and to remain in a ”normalized” problem. However, this simple solution does not guarantee efficiency. Instead, we propose to fit a distribution to each pair of log-likelihoods and later combine them to produce belief values. We set this hesitation distribution experimentally and eventually tune it on a validation set. We use a simple model which assumes the belief in the hesitation distribution to follow a zero-mean Gaussian function with standard deviation, σ , with respect to the margins (the differences of scores). We define σ as $\sigma = \sqrt{\alpha \cdot \sigma_s}$, where σ_s is the variance of the margins of pairwise log-likelihoods for the HMM case. The coefficient α controls the level of uncertainty in the belief function. The bigger it is, the more hesitation the belief function contains. If α is too

small, the belief function will be equivalent to a max function over the likelihoods: $\text{argmax}(\mathcal{L}(\cdot))$ will focus all the belief, and the rest will be zero-believed ($\mathcal{L}(\cdot)$ denotes the likelihood of each model). On the contrary, if α is too big, the belief is focused on the widest hypothesis (the complete hesitation), and making a non-random decision over such a function is impossible.

Having defined the hesitation distribution, the EBFs are calculated and they are fused together with Dempster's rule of combination:

$$m = \bigoplus_{i,j \neq i} m_{ij} \quad (12)$$

The entire algorithm to compute a belief function from a set of nonhomogeneous scores is given in Fig. 1. So far, we described how to obtain the hesitation distribution for each EBF, and the influence of these models on the global belief function (step 4 of Fig. 1), but we have not defined how to create this global belief function from the EBFs. This is done by (1) refining the EBFs so that they are defined on the complete powerset of Ω instead of on a part of it [22](see step 4 of Fig. 1), and (2) fusing all the EBFs with Dempster's rule of combination (step 5 of Fig. 1).

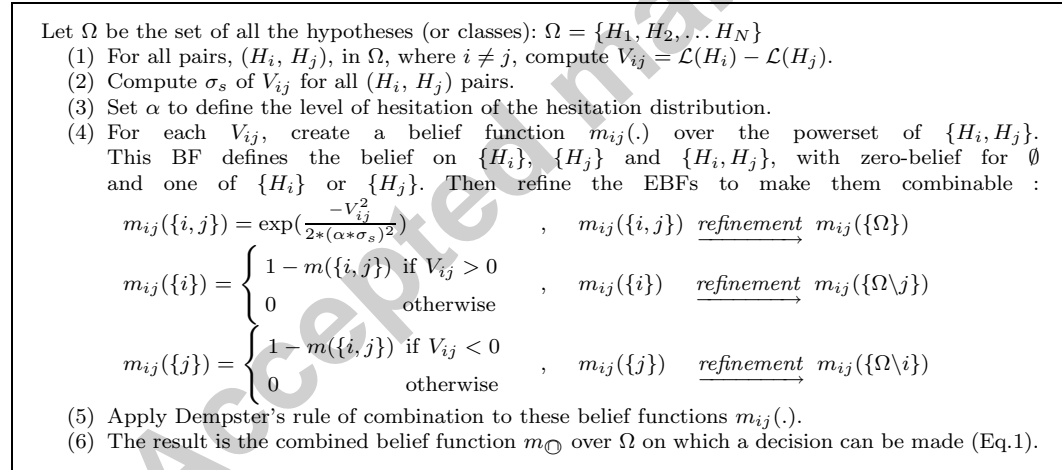


Fig. 1. Algorithm to compute beliefs from a set of nonhomogeneous scores.

4.2 Introducing uncertainty via Belief Functions

In our classification problem, we have signs with manual and non-manual information. We call signs that share the similar manual component as a “cluster”. We hypothesize that it will be easier to differentiate between signs from different clusters. The signs in the same cluster can be differentiated via the non-manual component. However, the non-manual features can be noisy as

a result of the data collection setup (i.e. 2D acquisition instead of 3D) and the feature extraction process. In some signs, the hand can be in front of the head, whereas in others, the face detector fails. Thus, we consider a potential absence of information with respect to this non-manual component. Then, our purpose is to make a decision only when it is sensible to do so, but also to accept a part of indecision when the information is not meaningful. As this is the purpose of the γ -PPT when $\gamma > 1$, we propose to apply the belief formalism to the problem, i.e:

- To model the hesitation among the gestures by a belief function computed from a set of likelihoods.
- To make a decision with the PPT (Eq. 4).

With respect to the quality of the information available among the features, we assume that the decision between the clusters will be complete, but a hesitation may remain within a cluster (concerning the non-manual component). In order to make a final decision within a cluster, we need a second stage of decision.

4.3 Sequential fusion with uncertainty

The sequential belief based fusion technique consists of two classification phases where the second is only applied when there is hesitation (see Fig. 2). 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 (via the PPT) is evaluated and resolved via the second bank of HMMs. In this setup, the assumption is that the HMMs of the first bank are more 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 in which the second bank of HMMs are capable of discriminating.

5 Methodology & Experiments

In order to assess the appropriateness of our belief-based method, we have performed experiments to compare it with several other mechanisms for fusing manual and non-manual signs. The experiments are conducted on a sign language database which has been collected during the eNTERFACE'06 workshop. In the following section, we give details about this database.

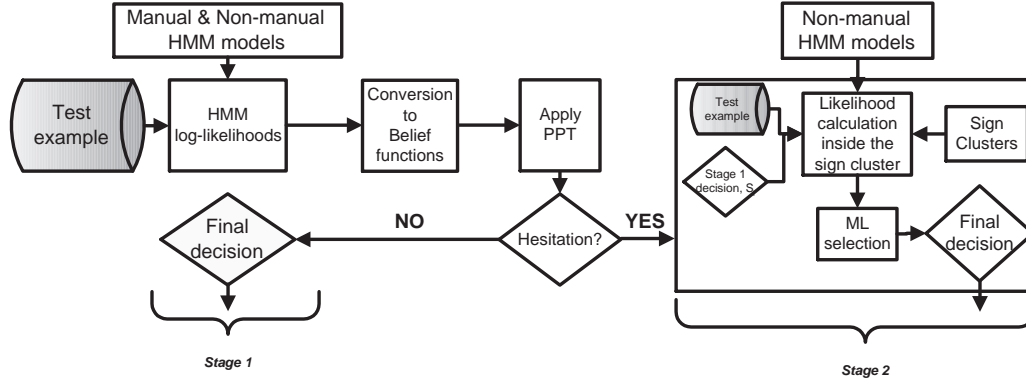


Fig. 2. Sequential belief-based fusion flowchart

Table 2
Signs in eNTERFACE'06 Database

Base Sign	Variant	Hand Motion Variation	Head Motion (NMS)
Clean	Clean		
	Very clean		✓
Afraid	Afraid		
	Very afraid	✓	✓
Fast	Fast		
	Very fast		✓
Drink	To drink		✓
	Drink (noun)	✓	
Open (door)	To open		
	door (noun)	✓	

Base Sign	Variant	Hand Motion Variation	Head Motion (NMS)
Here	[smbdy] is here		✓
	Is [smbdy] here?		✓
	[smbdy] is not here		✓
Study	Study		
	Study continuously	✓	✓
	Study regularly	✓	✓
Look at	Look at		
	Look at continuously	✓	✓
	Look at regularly	✓	✓

5.1 eNTERFACE'06 ASL Database

The signs in the eNTERFACE'06 ASL Database [23] 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 systematic variations of the base signs in the form of non-manual signs, or inflections in the signing of the same manual sign. A base sign and its variants will be called a “base sign cluster” for the rest of this paper. Table 2 lists the signs in the database. As observed from Table 2, some signs are differentiated only by the head motion; some only by hand motion variation and some by both. Two example signs are illustrated in Fig.3.

A single web camera with 640×480 resolution and 25 frames per second rate is used for the recordings. The camera is placed in front of the subject. The database is collected from eight subjects, each performing five repetitions of

each sign. The dataset is divided to training and test sets where 532 examples are used for training (28 examples per sign) and 228 examples for reporting the test results (12 examples per sign). The subjects in training and test sets are different except for one subject whose examples are divided between training and test sets. The distributions of sign classes are equal both in training and test sets. For the cases where a validation set is needed, we apply a stratified 7-fold cross validation (CV) on the training set. Since we concentrate on the



Fig. 3. Example signs from eINTERFACE'06 ASL database (a) HERE and NOT HERE, (b) CLEAN and VERY CLEAN

fusion step in this paper, we have directly used the processed data from [23] where the features of hand shape, hand motion and head motion are extracted. Sign features are extracted both for manual signs (hand motion, hand shape, hand position with respect to face) and non-manual signs (head motion). 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: x , y coordinates of CoM, and horizontal, vertical velocity, form the hand motion features. Hand shape features are appearance-based shape features calculated on the binary hand images. These features include the parameters of an ellipse fitted to the binary hand and statistics from a rectangular mask placed on top of the binary hand. For head motion analysis, the system detects rigid head motions such as head rotations and head nods [24]. The orientation and velocity information of the head and the quantity of motion are used as head motion features. Further details can be found in [23].

5.2 Clustering for Sequential Fusion

In this context, we define a sign cluster as a group of signs which are similar and the differences are either based on the non-manual component or variations of

the manual component. From the semantic interpretation of the signs in the database (see Table 2), we can define the base sign clusters as shown with the bold lines in Fig. 5. In Figures 5a and b the bold lines indicate the semantic clusters.

In a classification task, although one can utilize prior knowledge such as the sign clusters based on semantic information, this has some disadvantages. First, it is not guaranteed that these semantic clusters are suitable for the classification task, and second, the trained model will be database dependent and extending the database with new signs will require the re-definition of the cluster information. Thus, an automatic clustering method that depends on the data and considers the capabilities of the classifier would be preferable. We propose two methods for automatic identification of the clusters: The first method is based on belief formalism, and the second method is based on the classification errors without any belief calculation. For the latter, we propose to use the confusion matrix for cluster identification; and for the former we propose to use the hesitation matrix. In both cases, the cluster identification is done by applying 7-fold CV on the training data. The confusion/hesitation matrices of each fold are combined to create a joint matrix, which is used to identify the clusters (Fig. 4).

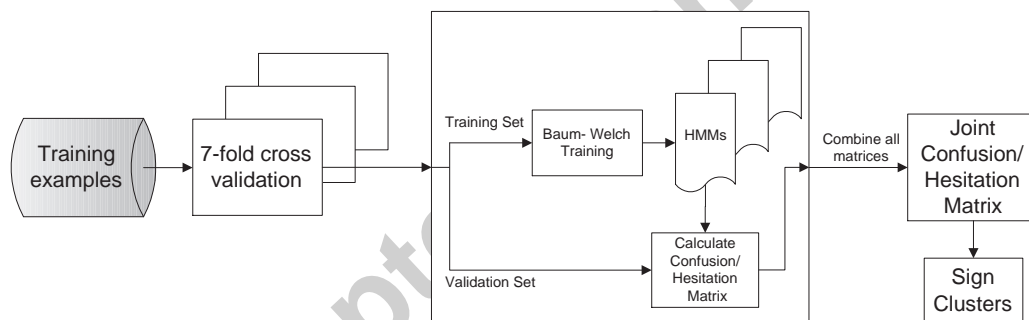


Fig. 4. Identifying sign clusters by cross validation via confusion or hesitation matrices

To convert a confusion matrix to a sign cluster matrix, for each sign, we cluster all signs among which there is confusion. For example, assume that sign i is only confused with sign j . Then the sign cluster of class i is (i, j) . The sign cluster of class j is separately calculated from its confusions in the estimation process. Fig. 5a shows the sign clusters identified via the confusion matrix for the eNTERFACE'06 sign data.

In belief formalism, we use the uncertainties in the decisions of the first stage classifier to define the sign clusters (Fig. 5b). For this purpose, only the elements which are hesitation-prone are considered in a hesitation matrix. The elements without any hesitation, either complete mistakes or correct ones are excluded from the calculation of this hesitation matrix. This is equivalent to a confusion matrix but the sum over the matrix is equal to the number of hesi-

tations. Each hesitation is multiplied by the number of elements among which the hesitation occurs. Then this matrix is transformed so that it is closed, transitive and reflexive. The classes of equivalence in this matrix correspond to the clusters. The number of elements within each hesitation is directly related to the γ parameter. Hence, the creation of the clusters is directly correlated with the tuning of the PPT on the validation set. To simply tune it, we propose to approximately set it to the number of signs within the base sign clusters (i.e. $\gamma = 2$ or 3).

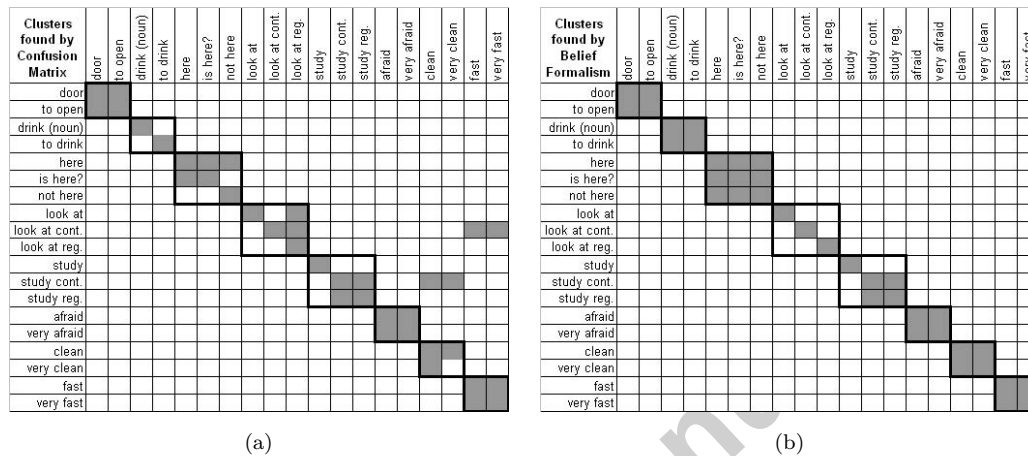


Fig. 5. Sign clusters (a) identified by the joint confusion matrix of 7-fold CV, (b) identified by the uncertainties between the classes in 7-fold CV. Clusters are shown row-wise, where for each sign row, the shaded blocks show the signs in its cluster. Bold lines show the clusters indicated by prior semantic knowledge, which we do not use.

Fig. 5 shows the sign clusters identified by the two techniques, by confusions in the joint confusion matrix and by uncertainties provided by the belief functions, respectively. Boldly outlined squares show the base sign clusters and for each row, shaded blocks show the identified clusters for the corresponding sign. The problem with the confusion matrix method is its sensitivity. Even a single mistake causes a cluster formation. On the other hand, the belief based method robustly identifies the uncertainties and provides robust clusters.

At this point it is helpful to discuss the clustering results and their interpretation with respect to the signs in the database. As listed in Table 2, there are eight base signs in the database. The 19 signs are formed either by adding non-manual information to the base sign or by variations in the signing of the base sign and sometimes, both. Thus in a sign cluster, not all the confusions can be resolved by utilizing only non-manual information. Here we give the details of the base signs and discuss the clustering results shown in Fig. 5b.

- Signs DOOR and TO OPEN are only differentiated by the positioning of the hands and their speed and there is no non-manual information to differ-



Fig. 6. (a) DRINK, and (b) TO DRINK differ with the head motion and variations in the hand motion

entiate between them. Although the clustering method puts these two signs in the same cluster, one can not expect to have a correct decision at the second step by only utilizing the non-manual component. For these signs the confusion must be resolved at the first step by the manual information.

- Signs DRINK and TO DRINK are differentiated by both non-manual information and variations in signing (See Fig. 6). TO DRINK sign imitates a single drinking action with the head motion. DRINK sign is performed without head motion and with repetitive hand motion. However, as the hand is in front of the mouth region, for some frames, the face detector fails and provides wrong feature values that mislead the recognizer.
- Signs HERE, IS HERE, NOT HERE have exactly the same manual sign but the non-manual sign differs. Thus when only manual information is used, confusions between these three signs are expectable. Non-manual information resolves the confusions in the cluster.
- For the LOOK AT sign, the differentiation is provided by both non-manual information and variations in signing. However, the hands can be in front of the head for many of the frames. For those frames, the face detector fails and provides wrong feature values that mislead the recognizer.
- The STUDY sign: It is interesting to observe that in Fig. 5b, the base study sign is clustered into two sub-clusters. This separation agrees with the nature of these signs: In the sign STUDY, there is a local finger motion without any global hand motion, and this directly differentiates this sign from the other two variations (See Fig. 7). The confusion between the manual components of STUDY REGULARLY and STUDY CONTINUOUSLY can stem from a deficiency of the 2D capture system. The hand motion of these two signs differ mainly in depth. However, the non-manual components can be used at the second stage to resolve this confusion.
- For signs AFRAID, FAST, and CLEAN, a non-manual sign is used to emphasize their meaning (signs VERY AFRAID, VERY FAST, and VERY CLEAN). Each of these signs and their emphasized versions are put in the same cluster and the confusion inside these clusters can be resolved by utilizing the non-manual component.



Fig. 7. (a) STUDY, and (b) STUDY REGULARLY differ with the head motion and global hand motion

5.3 Reference Algorithms

To model the manual and non-manual components of the signs and perform classification, we train three different HMMs:

- HMM_M : Uses only manual features (hand motion, shape and position with respect to face)
- HMM_N : Uses only non-manual features (head motion)
- $HMM_{M\&N}$: Uses both manual and non-manual features (manual features plus head motion)

5.3.1 HMM Classification and Feature Level Fusion

We train HMMs for each sign and classify a test example by selecting the sign class whose HMM has the maximum log-likelihood. The HMM models are selected as left-to-right 4-state HMMs with continuous observations where Gaussian distributions with full covariance are used to model the observations at each state. The Baum-Welch algorithm is used for HMM training. Initial parameters of transition, prior probabilities and initial parameters of Gaussians are selected randomly.

We compare the classification performance of HMM_M and $HMM_{M\&N}$ to see the information added by the non-manual features via feature level fusion. The classification results of these two models should show us the degree of effective utilization of the non-manual features when combined into a single feature vector with manual features. Although there is no direct synchronization between the manual and non-manual components, the second model, $HMM_{M\&N}$, models the dependency of the two components for sign identification. The classification results and confusion matrices are shown in Fig. 8. Although the classification accuracy of $HMM_{M\&N}$ is slightly better than HMM_M , total accuracy is still low. The high dimensionality of the feature vector (61 features per frame) can be a cause of this low accuracy. The curse of dimensionality affects HMM training. Another factor is that the non-manual features can be noisy as a result of wrong face detection, especially when hands are in front

Only Hand Features	door	to open	drink (noun)	to drink	here	is here?	not here	look at	look at cont.	look at reg.	study	study cont.	study reg.	afraid	very afraid	clean	very clean	fast	very fast	
door	10	2																		
to open	0	12																		
drink (noun)			12	0																
to drink			1	11																
here					4	3	5													
is here?					0	5	7													
not here					0	5	7													
look at								7	1	4										
look at cont.								0	12	0										
look at reg.	3	1						0	1	7										
study											4	4	4							
study cont.											0	8	4							
study reg.											1	1	10							
afraid														2	10					
very afraid														0	12					
clean																6	6			
very clean																2	10			
fast																			3	8
very fast										1									1	11

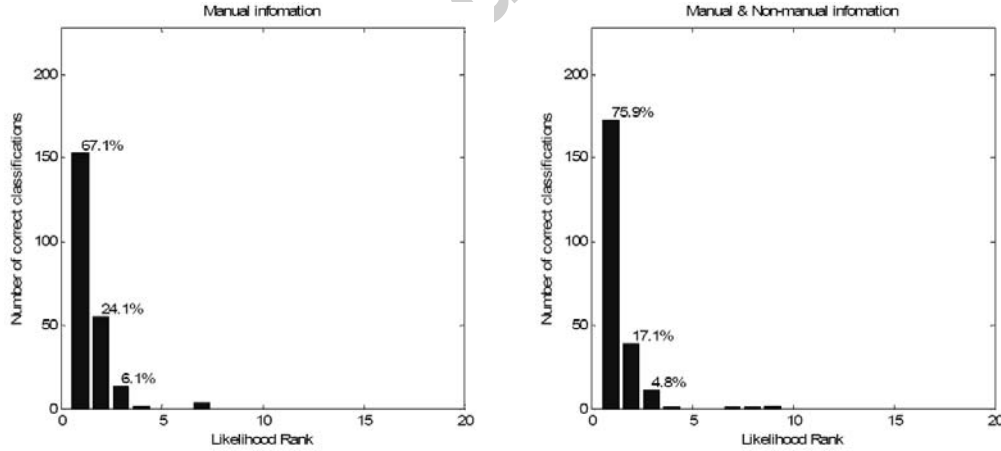
(a) 97.8% base sign accuracy, 67.1% total accuracy

Hand Head feature fusion	door	to open	drink (noun)	to drink	here	is here?	not here	look at	look at cont.	look at reg.	study	study cont.	study reg.	afraid	very afraid	clean	very clean	fast	very fast	
door	11	1																		
to open	1	11																		
drink (noun)			12	0																
to drink			0	12																
here					4	4	4													
is here?					0	12	0													
not here					0	2	10													
look at								7	1	4										
look at cont.								0	12	0										
look at reg.	1							0	4	7										
study											3	0	9							
study cont.											0	8	4							
study reg.											0	2	10							
afraid														3	9					
very afraid														0	12					
clean																11	1			
very clean																2	10			
fast																			6	6
very fast																			0	12

(b) 99.5% base sign accuracy, 75.9% total accuracy

Fig. 8. (a) Confusion matrix of HMM_M , (b) Confusion matrix of $HMM_{M\&N}$. Rows indicate the true class and columns indicate the estimated class. Base sign and its variations are grouped and shown in bold squares. In both of the cases, classification errors are mainly between variations of a base sign.

of the face. Besides, non-manual information is a secondary component of the sign and when analyzed together, the manual information may override the non-manual part. In any case, although it causes an improvement, non-manual information is not effectively utilized by feature fusion in HMM classification. However, it is worth noting that the classification errors in both of the models are mainly between variants of a base sign and out of cluster errors are very few. A further investigation of the classification results show that in about



(a) Rank 3 accuracy: 97.3%

(b) Rank 3 accuracy: 97.8%

Fig. 9. Rank distribution of the true class likelihoods of (a) HMM_M , (b) $HMM_{M\&N}$

97.5% of the examples, the true class resides among the first three highest likelihoods, if not the maximum (see Fig. 9). By further analyzing the first three highest likelihoods, one might increase the performance.

5.3.2 Parallel Score Level Fusion

Results of the previous section show that manual and non-manual information are not effectively utilized by feature level fusion. In parallel score level fusion, two independent experts are used and the scores (confidences, likelihoods . . .) of these experts are combined with several combination rules. In this setup, the idea is to use one expert that models manual information (HMM_M) or manual and non-manual information together ($HMM_{M\&N}$) and to combine the scores of this expert with another one that models non-manual information (HMM_N). We use the sum rule to combine the log-likelihoods of the HMM experts. To use as a reference fusion method, we applied parallel fusion of the scores of (1) HMM_M and HMM_N , (2) $HMM_{M\&N}$ and HMM_N . The comparative results are shown in Table 3.

5.4 Sequential Score Level Fusion

Our proposed belief-based sequential fusion mechanism aims to identify the cluster of the sign in the first step and to resolve the confusion inside the cluster in the second step. For comparison purposes, we propose another sequential fusion technique which follows the same strategy but only uses the HMM likelihoods without any belief formalism. In each of these techniques, information related to the sign cluster must be provided. In this section, we summarize each sequential fusion method.

5.4.1 Sequential Fusion based on HMM Likelihoods

In this method, the fusion methodology is based on the likelihoods of the HMMs. The sign clusters are automatically identified during training by the joint confusion matrix of 7-fold CV (see Fig. 5a). The first step selects the sign with the maximum likelihood of $HMM_{M\&N}$. The cluster of the selected sign is determined by the previously identified clusters. In the second stage the classifier only considers signs within the identified cluster. The decision is made by selecting the sign with the maximum likelihood of HMM_N .

5.4.2 Sequential Fusion based on Belief Functions and Uncertainties

A two step sequential fusion technique in which the sign clusters are automatically identified during training by the uncertainties calculated from the beliefs on each sign is utilized (see Fig. 5). The difference of this method from the likelihood-based one is twofold: (1) the cluster identification method is based on belief functions (2) It is not mandatory to proceed to the second stage. If our belief about the decision of the first stage is certain, then we use that

decision. The steps of the technique can be summarized as follows (also see Fig. 2):

- (1) Automatically identify the clusters via uncertainties.
- (2) Define the parameter γ for PPT and the parameter σ for the hesitation pattern.
- (3) Convert the likelihoods of the first stage HMMs to belief functions as explained in Section 4.1.
- (4) Apply the PPT.
 - (a) If there is no uncertainty in the result, decide accordingly.
 - (b) Otherwise, identify the cluster of this sign and proceed to the second stage.
- (5) In the second stage, only consider the signs within the identified cluster.
- (6) The decision is made by selecting the sign with the maximum likelihood of HMM_N .

5.5 Results

The accuracies with different fusion techniques are summarized in Table 3. The automatically identified clusters using 7-fold CV can be seen in Figs. 5a and 5b, for the two techniques, fusion using likelihoods and fusion using belief functions, respectively. We have used the notations \Rightarrow and \rightarrow , respectively for these two fusion methods to indicate the difference in the process of proceeding to the second stage, where \Rightarrow indicates unconditional proceeding whereas \rightarrow indicates a condition based on the belief-based analysis. Base sign clusters are as defined in the previous sections. Manually defined sign clusters are tuned manually by the human expert to emphasize the fact that even if it is tuned manually by taking the properties of the classification and analysis methods into consideration, the proposed method with automatically defined clusters is superior. Sequential-belief based fusion has the highest accuracy (81.6%) among all implemented techniques. The reason for this success is both based on the robustness of the belief-based cluster identification and the possibility of accepting the first stage classification decision thanks to belief formalism. The effect of the latter can be seen from the last two lines of Table 3, where the same clustering result, based on uncertainties calculated from belief functions, gives a very low accuracy if we do not apply belief-based decision analysis. For the former, when the clusters are not properly and robustly defined, the classification performance may degrade. This effect can be seen in Table 3 where we report the accuracies of sequential likelihood fusion with different cluster identification techniques. When compared with the base model, $\text{HMM}_{M\&N}$, the classification accuracy is lower in three of the cluster identification methods and only higher with manually defined clusters. The main reason is that when belief-based decision analysis is not applied,

Table 3
Classification performance.

	Models	Fusion method	Cluster identification	Test Accuracy
Reference	HMM_M	No fusion	-	67.1%
	$HMM_{M\&N}$	Feature	-	75.9%
	$HMM_M + HMM_N$	Parallel	-	70.6%
	$HMM_{M\&N} + HMM_N$	Parallel	-	78.1%
Proposed	$HMM_{M\&N} \Rightarrow HMM_N$	Sequential likelihood	Base sign clusters	73.7%
	$HMM_{M\&N} \Rightarrow HMM_N$	Sequential likelihood	Manually defined	78.1%
	$HMM_{M\&N} \Rightarrow HMM_N$	Sequential likelihood	Automatic via confusion matrix	75.0%
	$HMM_{M\&N} \Rightarrow HMM_N$	Sequential likelihood	Automatic via uncertainties	73.3%
	$HMM_{M\&N} \rightarrow HMM_N$	Sequential belief-based	Automatic via uncertainties	81.6%

it proceeds to the next stage unconditionally, regardless of the first stage decision, and correct classifications of the first step are altered.

Although the time dependency and synchronization of manual and non-manual features are not that high, feature fusion still improves the classification performance (13% improvement) by providing extra features of non-manual information. This improvement is also superior to parallel score fusion of manual and non-manual models, showing the need for co-modeling. However, the modeling of $HMM_{M\&N}$ is not sufficient and still has low accuracy. The classification performance is improved by adding an extra expert, HMM_N , and by performing parallel score level fusion with another expert, $HMM_{M\&N}$ (3% improvement to $HMM_{M\&N}$ and 16% improvement to HMM_M). However, the sequential belief-based fusion and the clustering idea is superior since the manual information forms the primary component and the non-manual information forms the secondary component of a sign. The sequential belief-based fusion method processes the signs according to this information. The analysis of wrongly classified examples shows that in most of the cases the wrong decision is due to the lack of 3D information, or due to failure in face detection. Since we use 2D features, some of the signs resemble each other (e.g. signs STUDY CONTINUOUSLY and STUDY REGULARLY). Failures in face detection mainly occur when the hand occludes the face; thus resulting in wrong head tracking results. There are also a few examples where the wrong classifications are due to hand tracking mistakes.

6 Conclusions

We have compared several fusion techniques for integrating manual and non-manual signs in a sign language recognition system. A dedicated fusion methodology is needed to cover the specialties of the usage of manual and non-manual signs in sign languages. We propose to use a two-step fusion methodology: The first step mainly depends on the manual information and the second step only utilizes non-manual information. This is inspired by the fact that the manual component is the main component in sign language communication. In many of the signs, the information can be conveyed without the need for non-manual signals. However, when a non-manual sign is used, it may radically change the meaning of the manual sign, by either emphasizing it, or indicating a variation. Hence, one can not discard the non-manual signs in a complete recognition system. Our proposed belief-based fusion mechanism is based on this observation and applies a two-step fusion approach. The first step of the fusion mechanism applies feature level fusion on manual and non-manual components and attempts to make a decision. The decision is analyzed by the belief formalism and by considering a potential absence of information, and it can be accepted or a hesitation between some of the sign classes can be expressed. This hesitation is expected to remain inside the sign cluster and the second step aims to resolve the hesitation by considering only the non-manual component.

The key point of our belief-based fusion approach is two-fold. First, it has a two-step decision phase and if the decision at the first step is without hesitation, the decision is immediately made, without proceeding to the next step. This would speed up the system, since there is no need for further analysis. Even in the case of a hesitation, the decision of the first step identifies the cluster which the test sign belongs to, if not the exact sign class. Second, 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, then training models for the new signs and running the belief formalism to find the new sign clusters.

These two key points root in the capability of the PPT to make a decision which is a balance between the risk of a complete decision and the cautiousness of a partial decision. It is able to provide a singleton decision when supported, but on the other hand, as long as the information is too hesitation-prone, it makes an incomplete decision. Then, it is automatically decided whether the second stage is used or not. Our results show that automatic belief based clustering even outperforms manual labeling based on semantic information in identifying base sign clusters and variations within clusters. Finally, this methodology can also be used in other linked problems, such as identifying grammatical processes or performance differences in sign languages provided that necessary features are extracted.

7 Acknowledgments

This work is a result of a cooperation supported by SIMILAR 6FP European Network of Excellence (www.similar.cc). This work has also been supported by TUBITAK project 107E021 and Bogazici University project BAP-03S106.

References

- [1] W. C. Stokoe, Sign language structure: An outline of the visual communication systems of the American deaf, *Studies in Linguistics: Occasional papers* 8.
- [2] S. K. Liddell, *Grammar, Gesture, and Meaning in American Sign Language*, Cambridge University Press, 2003.
- [3] Y. Wu, T. S. Huang, Hand modeling, analysis, and recognition for vision based human computer interaction, *IEEE Signal Processing Magazine* 21 (1) (2001) 51–60.
- [4] S. C. W. Ong, S. Ranganath, Automatic sign language analysis: A survey and the future beyond lexical meaning., *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27 (6) (2005) 873–891.
- [5] Q. Munib, M. Habeeba, B. Takruria, H. A. Al-Malika, American sign language (ASL) recognition based on Hough transform and neural networks, *Expert Systems with Applications* 32 (1) (2007) 24–37.
- [6] N. Liu, B. C. Lovell, P. J. Kootsookos, R. I. A. Davis, Model structure selection and training algorithms for an HMM gesture recognition system, in: *Ninth International Workshop on Frontiers in Handwriting Recognition (IWFHR'04)*, IEEE Computer Society, Washington, DC, USA, 2004, pp. 100–105.
- [7] C. Vogler, D. Metaxas, Parallel hidden Markov models for American sign language recognition, in: *International Conference on Computer Vision, Kerkyra, Greece, Vol. 1, 1999*, pp. 116–122.
- [8] K. Ming, S. Ranganath, Representations for facial expressions, in: *International Conference on Control Automation, Robotics and Vision, Vol. 2, 2002*, pp. 716–721.
- [9] U. Erdem, S. Sclaroff, Automatic detection of relevant head gestures in American sign language communication, in: *International Conference on Pattern Recognition, Vol. 1, 2002*, pp. 460–463.
- [10] J. Ma, W. Gao, R. Wang, A parallel multistream model for integration of sign language recognition and lip motion, in: *Third International Conference on Advances in Multimodal Interfaces (ICMI '00)*, Springer-Verlag, London, UK, 2000, pp. 572–581.

- [11] A. P. Dempster, A generalization of Bayesian inference, *Journal of the Royal Statistical Society. Series B (Methodological)* 30 (2) (1968) 205–247.
- [12] G. Shafer, *A Mathematical Theory of Evidence*, Princeton University Press, 1976.
- [13] P. Smets, R. Kennes, The transferable belief model, *Artificial Intelligence* 66 (2) (1994) 191–234.
- [14] J. Kohlas, P. A. Monney, *Theory of evidence: A survey of its mathematical foundations, applications and computations*, *ZOR-Mathematical Methods of Operational Research* 39 (1994) 35–68.
- [15] B. Cobb, P. Shenoy, A comparison of methods for transforming belief functions models to probability models, *Lecture Notes in Artificial Intelligence* 2711 (2003) 255–266.
- [16] M. Daniel, Probabilistic transformations of belief functions, in: *ECSQARU*, 2005, pp. 539–551.
- [17] P. Smets, Decision making in a context where uncertainty is represented by belief functions, *Physica-Verlag, Heidelberg, Germany, IRIDIA*, 2002, Ch. *Belief Functions in Business Decisions*, pp. 17–61.
- [18] G. Shafer, P. P. Shenoy, Local computation in hypertrees, *Tech. rep.*, School of Business, University of Kansas (1991).
- [19] B. R. Cobb, P. P. Shenoy, On the plausibility transformation method for translating belief function models to probability models., *International Journal of Approximate Reasoning* 41 (3) (2006) 314–330.
- [20] L. R. Rabiner, A tutorial on hidden Markov models and selected applications in speech recognition, in: *Proceedings of the IEEE*, Vol. 77 of no. 2, 1989, pp. 257–285.
- [21] O. Aran, T. Burger, A. Caplier, L. Akarun, Sequential belief based fusion of manual and non-manual signs, in: *Gesture Workshop*, 2007.
- [22] T. Burger, O. Aran, A. Caplier, Modeling hesitation and conflict: A belief-based approach for multi-class problems, in: *ICMLA '06: Proceedings of the 5th International Conference on Machine Learning and Applications*, IEEE Computer Society, Washington, DC, USA, 2006, pp. 95–100.
- [23] O. Aran, I. Ari, A. Benoit, P. Campr, A. H. Carrillo, F.-X. Fanard, L. Akarun, A. Caplier, M. Rombaut, B. Sankur, Signtutor: An interactive system for sign language tutoring, *IEEE Multimedia*, to appear.
- [24] A. Benoit, A. Caplier, Head nods analysis: Interpretation of non verbal communication gestures, in: *International Conference on Image Processing, ICIP2005, Genova, Italy, Vol. 3*, 2005, pp. 425–428.

OYA ARAN received the B.S. and M.S. degrees in Computer Engineering from Bogazici University, Istanbul, Turkey in 2000 and 2002, respectively. She is currently doing her PhD at Bogazici University. Her current research interests include sign language recognition, human-computer interaction, and machine learning.

Accepted manuscript

Thomas Burger graduated from the Institut National Polytechnique de Grenoble in telecom engineering in 2004, meanwhile receiving a MS degree in Combinatorial Optimization. At present he is concluding a PhD in gesture recognition at France Telecom R&D.

Accepted manuscript

ALICE CAPLIER is graduated from the École Nationale Supérieure des Ingénieurs Électriciens de Grenoble (ENSIEG) of the Institut National Polytechnique de Grenoble (INPG), France, in 1991. She obtained her Master's degree in Signal, Image, Speech Processing and Telecommunications in 1992 and her PhD from the INPG in 1995. Since 1997 she is teaching at the École Nationale Supérieure d'Électronique et de Radio électricité de Grenoble (ENSERG) of the INPG and is a permanent researcher at the Laboratoire des Images et des Signaux (LIS) in Grenoble. Her interest is on human motion analysis and interpretation.

Accepted manuscript

LALE AKARUN received the B.S. and M.S. degrees in electrical engineering from Bođaziçi University, Istanbul, Turkey, in 1984 and 1986, respectively, and the Ph.D. degree from Polytechnic University, Brooklyn, NY, in 1992. From 1993 to 1995, she was Assistant Professor of electrical engineering at Bođaziçi University, where she is now Professor of computer engineering. Her current research interests are in image processing, computer vision, and computer graphics.

Accepted manuscript