Available online at www.sciencedirect.com
ScienceDirect
Speech Communication 51 (2009) 820–830
SPEECH COMMUNICATION
www.elsevier.com/locate/specom

Maximum likelihood linear programming data fusion for speaker recognition

Enric Monte-Moreno [a,*], Mohamed Chetouani [b], Marcos Faundez-Zanuy [c], Jordi Sole-Casals [d]

[a] *TALP Research Center, UPC Barcelona, Spain*
[b] *Université Pierre et Marie Curie-Paris 6, France*
[c] *Escola Univesitària Politècnica de Mataró, UPC Barcelona, Spain*
[d] *Universitat de Vic, Barcelona, Spain*

Received 11 January 2008; received in revised form 16 April 2008; accepted 21 May 2008

Abstract

Biometric system performance can be improved by means of data fusion. Several kinds of information can be fused in order to obtain a more accurate classification (identification or verification) of an input sample. In this paper we present a method for computing the weights in a weighted sum fusion for score combinations, by means of a likelihood model. The maximum likelihood estimation is set as a linear programming problem. The scores are derived from a GMM classifier working on different feature extraction techniques. Our experimental results assessed the robustness of the system in front changes on time (different sessions) and robustness in front of changes of microphone. The improvements obtained were significantly better (error bars of two standard deviations) than a uniform weighted sum or a uniform weighted product or the best single classifier. The proposed method scales computationally with the number of scores to be fusioned as the simplex method for linear programming.

Keywords: Speaker recognition; Data fusion; GMM; Maximum likelihood; Linear programming; Simplex

1. Introduction

Biometric recognition (Faundez-Zanuy, 2006) offers a promising approach for security applications, with some advantages over the classical methods, which depend on something you have (key, card, etc.), or something you know (password, PIN, etc.). A nice property of biometric traits is that they are based on something you are or something you do, so you do not need to remember anything neither to hold any token.

On the other hand, they have an important drawback, because if a person's biometric data is stolen, it is not possible to replace it (Faundez-Zanuy, 2004). Probably, these drawbacks have slowed down the spread of use of biometric recognition (Faundez-Zanuy, 2005b). For those applications with a human supervisor (such as border entrance control), this can be a minor problem, because the operator can check if the presented biometric trait is original or fake. However, for remote applications such as internet, some kind of liveliness detection and anti-replay attack mechanisms should be provided. Fortunately, speech offers a richer and wider range of possibilities when compared with other biometric traits, such as fingerprint, iris, hand geometry, face, etc. For instance, you can use a text-dependent system (Faundez-Zanuy and Monte-Moreno, 2005) and to ask the user for a specific speech sentence. Speaker recognition does not offer the same robustness and precision than other biometric traits such as fingerprint and iris. However, strong efforts are done to enhance the performance, due to its particular set of characteristics that can permit to manage some vulnerability attacks.

* Corresponding author.
E-mail address: enric@gps.tsc.upc.edu (E. Monte-Moreno).

0167-6393/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.specom.2008.05.009

This paper is organized as follows: Section 2 describes the different data levels for fusion with special emphasis on the score level. A new strategy is presented for data fusion. Section 3 is devoted to the experimental results, and Section 4 summarizes the main conclusions.

2. Data fusion

2.1. Introduction

Given a biometric system, such as that depicted in Fig. 1, four main data fusion levels can be defined: sensor, feature, score (also known as opinion) and decision. The description of these levels is beyond the scope of this paper and can be found in (Faundez-Zanuy, 2005a).

In this paper we will focus on the score level. This kind of fusion is also known as confidence level. Given a set of classifiers (matchers), it consists of the combination of the scores provided by each matcher. The matcher just provides a distance measure or a similarity measure between the input features and the models stored on the database.

It is possible to combine several classifiers working with the same biometric characteristic (unimodal systems) or to combine different ones. In our case, it will be a unimodal combination, where both classifiers share the same input signal, as depicted in Fig. 2. This scheme can be easily generalized for more than two matchers.

2.2. Combination strategies

The score combination schemes for a given speaker can be done in several ways (see Kuncheva, 2004). The most natural strategies for combining different scores, might be

(1) Weighted sum: $O_s = \sum_{j=1}^{N} h_j o_{js}$.
(2) Weighted product: $O_s = \prod_{j=1}^{N} (o_{js})^{h_j}$.

In this paper we propose a fusion method, where the scores will be interpreted as probabilities of an observation, given a model. For each observation we will have a vector of N-scores, which will be the probability of the identity of a speaker obtained from a set of N classifiers. The global likelihood function will be the product of the all the probabilities (scores) of all speakers where each score (O_s) will be weighted by a factor h_j that will be specific for that score. The likelihood function of these probabilities can be understood as a fusion of either a weighted product of probabilities, or a weighted sum of logarithms of probabilities.

The estimation of the h_j parameters that weight the different scores can be done by several methods. The first and most simple might be the brute force method, which would consist on exploring the space of possible recognition rates for all possible combinations of a set of discrete values of the weighting parameters. The problem with this method

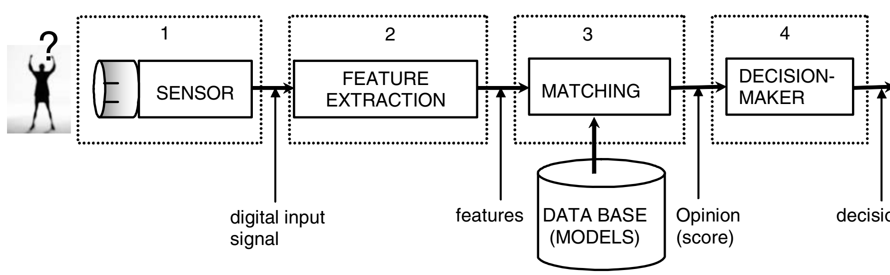

Fig. 1. General scheme of a biometric system.

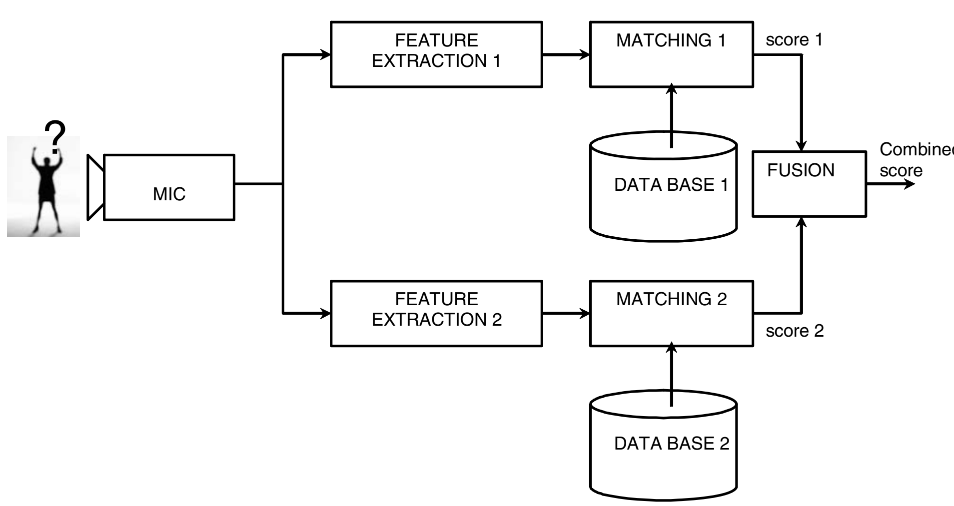

Fig. 2. General scheme for data fusion at score level.