Computers and Electrical Engineering 50 (2016) 10–25

Contents lists available at ScienceDirect

Computers and Electrical Engineering

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

Efficient online and offline template update mechanisms for speaker recognition☆

Anzar S.M.[a,*], Amala K.[a], Remya Rajendran[a], Ashwin Mohan[a], Ajeesh P.S.[a], Mohammed Sabeeh K.[a], Febin Aziz[b]

[a] Department of Applied Electronics and Instrumentation Engineering, M.E.S. College of Engineering, Kuttippuram, Kerala 679573, India
[b] Department of MCA, M.E.S. College of Engineering, Kuttippuram, Kerala 679573, India

ARTICLE INFO

Article history:
Received 21 April 2014
Revised 6 December 2015
Accepted 7 December 2015
Available online 29 December 2015

Keywords:
Template update
Mel frequency Cepstral coefficient
MFCC super template
GMM super model
Online update
Offline update

ABSTRACT

Sample variations are one of the main problems associated with speaker recognition. Most approaches use multiple templates in the gallery database. But, this requires enormous memory space. In order to minimize classification errors and intra-class variations, adaptive online and offline template update methods using vector quantization (VQ) and Gaussian mixture model (GMM) are proposed. Online and offline feature update as well as model update techniques are considered here. Feature update utilizes the vector quantization approach, while Gaussian mixture model approach is considered for model updating. The proposed methods automatically update the feature (model) in accordance with the biometric sample variations over time and they continually adapt the templates (user model) based on semi-supervised learning strategies. Experiments with 50 subjects reveal that the proposed template update strategies, improve the recognition accuracy and reduce the classification errors for voice recognition systems, even under sample variations.

1. Introduction

Need of security is increasing all over the world at every level: countries, societies, corporations and individuals. Biometrics play an important role to overcome the problem of security. Even though the human recognition task seems to be easy and straightforward, automated recognition becomes challenging and difficult. This is primarily due to the inherent variations in the acquisition process, undesired variations in the biometric data and non-ideal operating conditions such as background noise and non-uniform illumination [1]. Moreover, noisy data, intra-class variation, inter-class similarities, non-universality, spoofing, etc., are problems imposed by the biometric systems that tend to increase false accept rate (FAR) and false reject rate (FRR), ultimately reflecting towards poor performance. The new challenge for biometrics arises from trying to circumvent these limitations and developing more robust recognition systems [2]. The core of the biometric system is the 'template', which is a set of features (model) extracted from the biometric samples collected during the enrollment phase and stored in the system's database with a label related to the subject's identity. The success or failure of a biometric system heavily relies on the representativeness of the templates [3,4]. Template representativeness is a crucial problem with biometrics, as the input biometric data is subjected to

☆ Reviews processed and approved for publication by the Dr. M. Malek.

* Corresponding author. Tel.: +91 9447244119 (mobile); fax: +91 494 2698081.
E-mail addresses: anzar.sm.2014@ieee.org (Anzar S.M.), amalaramachandran@gmail.com (Amala K.), ramya.rajendran18@gmail.com (Remya Rajendran), ashwin9112@gmail.com (Ashwin Mohan), ajeesh20061992@gmail.com (Ajeesh P.S.), sabeehkmohammed@gmail.com (Mohammed Sabeeh K.), febifebin@gmail.com (Febin Aziz).

http://dx.doi.org/10.1016/j.compeleceng.2015.12.003
0045-7906/

on-going changes due to the presence of intra-class variations thus making the initially enrolled templates non-representative of them [5]. As a result, performance degradation arises. Trying to cover these intra-class variations by performing several enrollment sessions is expensive. Template update methods have been proposed in literature in order to account for feature variations due to time, aging, physiological and environmental factors [3,4]. The template update mechanism can automatically update the feature (model) in accordance with the biometric sample variance over time. By doing so, the feature/model in the database becomes an up-to-date representation of the user.

Various approaches are reported in the literature for improving the performance of biometric systems under sample variations[6]. Most of the state-of-the-art commercial systems employ fingerprints, face and voice biometrics. Compared to other modalities, speaker recognition systems have large public acceptance as they are easier to implement and the cost of the acquisition device is very cheap. One of the main advantages of voice biometrics is that, it is the only biometric that can be used for remote identification [2]. The greatest challenge in the design of voice based recognition system is to maintain high accuracy even under intra-class variations. Though a lot of work has been attempted on speaker adaptation [7], to the best of authors' knowledge, no work is reported for speaker template updating. A little attention was devoted to template update systems in research settings and academic publications, and this topic is not in the current mainstream of basic research in biometrics. We believe that, among the various reasons, the scarcity of appropriate databases, containing a sufficient number of biometric data collected over the time, and the intrinsic difficulty of this topic, also due to the lack of a precise formulation of the problem, hindered the advancement of this research field [4]. Hence, an attempt is made in this direction to address the same.

1.1. Major contributions

The primary objective of this work is to develop an adaptive template update mechanism for speaker recognition system, that can efficiently capture the intra-class variations of the user. The system should be capable of improving the recognition accuracy and robustness while reducing the false accept rate and false reject rate, even under sample variations. The performance of the proposed system is analysed in terms of recognition accuracy, classification errors and score density plots. The original contributions of the work are:

1. *Creation of a new databases for speaker update*: A new database for speaker update; English language database for adaptive speaker recognition (ELDASR) has been developed. The database contains 20 samples of 50 speakers which include both male and female voices. All the speakers are of age limit 20–23 years. The voice samples are collected in uncontrolled environmental settings in order to accommodate for the intra-class variations associated with the users.
2. *'MFCC super template' for speaker recognition*: One of the major steps behind template updating is the creation of an 'MFCC super template', formed by concatenating the MFCC features of each speaker samples of the individual speaker. The 'super template' acts as a representative vector for each speaker.
3. *Online and offline MFCC feature and GMM based model update*: Online and offline template update techniques based on 'self-update' is successfully implemented for both features as well as model updating.
4. *Secondary template (model) for speaker template (model) update*: Adaptive offline template (model) update strategy is implemented for speaker recognition system by introducing an additional secondary template (model) in conjunction with the primary template (model) created during the training phase. Secondary template update method is considered for both feature and model updating.

1.2. Organization of the paper

The rest of the paper is structured as follows. The following section gives a brief survey of the related works available in the literature. Section 3 gives the background of the work; i.e. brief outline of the speaker recognition system and the matching strategies adopted. Proposed online and offline techniques for speaker update is detailed in Section 4. Section 5 describes the database and experimental setup. The results of the experiments are detailed in Section 6. The paper concludes with a brief conclusion in Section 7.

2. Related work

The basic steps in template update methods are the assignment of identity labels. Identity label assignment can be done either in completely supervised (i.e. by human expert intervention) manner or by using automatic 'learning' (semi-supervised) methodology. The key difference between supervised and semi-supervised learning is the technique followed for the data labelling [3]. In supervised template update methods, the label assignment is manual, while in semi-supervised methods it is automatic. Supervised methods proposed so far are usually offline as they operate on a batch of collected samples. On the other hand, semi-supervised methods are automated methods that assign identity labels to the unlabelled data on the basis of their knowledge, derived through current enrolled templates without the intervention of human supervisor. These methods avoid the cost related to manual assignment of labels [3].

Uludag et al. proposed two methods to select a gallery of representative templates from multiple impressions collected at enrollment [8]. One of the methods was based on a clustering strategy, to choose a template set that best represents the intra-class variations, while the other selected templates that exhibit maximum similarity with the rest of the impressions. Methods