Information Fusion 32 (2016) 40–51

Contents lists available at ScienceDirect

Information Fusion

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

DPD-DFF: A dual phase distributed scheme with double fingerprint fusion for fast and accurate identification in large databases

Daniel Peralta[a,*], Isaac Triguero[b,c], Salvador García[a,d], Francisco Herrera[a], Jose M. Benitez[a]

[a] Department of Computer Science and Artificial Intelligence, CITIC-UGR (Research Center on Information and Communications Technology), University of Granada, 18071 Granada, Spain
[b] Department of Respiratory Medicine, Ghent University, 9000 Gent, Belgium
[c] Data Mining and Modelling for Biomedicine group, VIB Inflammation Research Center, 9052 Zwijnaarde, Belgium
[d] Department of Information Systems, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia

ARTICLE INFO

Article history:
Received 17 June 2015
Revised 7 February 2016
Accepted 13 March 2016
Available online 22 March 2016

Keywords:
Real-time identification
Large databases
Minutiae matching
Fingerprint fusion
Decision fusion
Score fusion
Parallel computing
Biometrics

ABSTRACT

Nowadays, many companies and institutions need fast and reliable identification systems that are able to deal with very large databases. Fingerprints are among the most used biometric traits for identification. In the current literature there are fingerprint matching algorithms that are focused on efficiency, whilst others are based on accuracy. In this paper we propose a flexible dual phase identification method, called DPD-DFF, that combines two fingers and two matchers within a hybrid fusion scheme to obtain both fast and accurate results. Different alternatives are designed to find a trade-off between runtime and accuracy that can be further tuned with a single parameter. The experiments show that DPD-DFF obtains very competitive results in comparison with the state-of-the-art score fusion techniques, especially when dealing with large databases or impostor fingerprints.

1. Introduction

Personal identification has arisen as an important issue in the last few years for many companies and institutions [1]. Identification databases grow larger every year, ranging from a few tens of people for small companies to several millions for institutions such as the police. Although there are various biometric traits that allow for identification, fingerprints are widely used because of their uniqueness and universality, among other properties [2,3]. Fingerprint recognition can be tackled from two different perspectives: verification [4] and identification [5]. The former consists of matching two fingerprints to determine whether they belong to the same finger or not. The latter aims to identify an input fingerprint from a set of fingerprints and determine which of them matches with the input. In this context, an Automatic Fingerprint Identification System (AFIS) is a tool that allows us to perform identifications in fingerprint databases [3].

Fingerprints are composed of a pattern of ridges and valleys, from which diverse features can be extracted. Among these features, minutiae are widely used for fingerprint matching, mostly due to their distinctiveness [2,6]. When two fingerprints are to be compared, the minutiae are extracted from the images, and then a matching algorithm is applied over the two minutiae sets to determine a similarity level. There are multiple proposals of minutiae-based matching algorithms in the literature [7]. Some of them are very efficient due to their simplicity [8], while others are very accurate [9]. However, these two objectives are usually not reached together because accurate algorithms tend to be complex, and therefore time-consuming. This restriction complicates the development of AFIS that are able to identify people in very large databases in a suitable time frame without precision loss. Moreover, as the overall response time of an identification procedure is linear with respect to the size of the database, even the fastest matching algorithms may become useless when the database grows too large. Moreover, the huge number of matchings causes an accuracy loss.

Information fusion is a widely used paradigm that improves overall precision in many fields, including biometrics [10–12]. In particular, two main approaches have been proven to enhance the recognition capabilities: the use of several fingerprint images

[*] Corresponding author. Tel.: +34 958244019; fax: +34 958243317.
E-mail addresses: dperalta@decsai.ugr.es (D. Peralta), Isaac.Triguero@irc.vib-UGent.be (I. Triguero), salvagl@decsai.ugr.es (S. García), herrera@decsai.ugr.es (F. Herrera), J.M.Benitez@decsai.ugr.es (J.M. Benitez).

http://dx.doi.org/10.1016/j.inffus.2016.03.002
1566-2535/© 2016 Elsevier B.V. All rights reserved.

[13] and the use of several matching algorithms [14]. The information fusion can be performed at different levels:

- Feature fusion approaches merge the characteristics extracted from different fingerprint images, coming either from the same finger or different fingers [15,16].
- Score fusion methods perform separate matchings and then sum up the scores [14,17].
- Decision fusion methods apply the matching algorithms in a hierarchical mode over the fingerprints [11,18].

Although these approaches increase the accuracy of the AFIS, they also slow the identification down because the processing workload is higher. In this work, we combine the ideas of multi-finger and multi-algorithm identification to improve the runtime along with the accuracy.

High Performance Computing (HPC) is an important tool to speed up the runtime of a system [19,20], and several proposals in the literature apply it to AFIS. However, these systems focus on objectives other than precision, such as high availability [21], load balancing [22] or reduced matching times [18]. Other systems provide the ability to identify in very large databases [23–25], but their accuracy is not improved with respect to a sequential AFIS.

There are currently several systems in the world that maintain large fingerprint databases. For example, as of September 2015, India's UIDAI system [26] stores the fingerprints of around 907 million people, although so far they are only used for verification purposes, not identification. FBI IAFIS [27] (now included within Next Generation Identification, NGI) keeps the fingerprints (among other data) for around 104 million subjects, and is able to perform searches in an average time of 72 minutes.

In this paper, we propose a flexible, Dual Phase Distributed AFIS with Double Fingerprint Fusion (called DPD-DFF) that integrates two fingerprints and two matching algorithms, aiming to overcome the weaknesses of isolated approaches: high identification time and accuracy loss. To do so, the identification is split into two phases, each of which can either use a single fingerprint or fuse two of them, conforming a mixed score fusion and decision fusion process:

- In the first phase, the database is explored by a fast matching algorithm to select a candidate set. Jiang's algorithm [8] has been selected for this phase due to its high running speed [7].
- Then, the second phase applies a more accurate algorithm to identify the correct identity within this candidate set. The matcher used in this phase is Minutia Cylinder-Code (MCC) [9], which is very precise [7].

With this design, the fingerprint fusion is powerful and flexible as it is performed at two separate levels. Furthermore, this strategy has been integrated within the parallel framework proposed in [23] in order to reach full scalability for arbitrarily large databases.

This manuscript is structured as follows. First, Section 2 provides the background information on the problem at hand. Section 3 presents DPD-DFF, the approach proposed in this paper. Section 4 describes the experiments performed and their results. Finally, Section 5 details the conclusions. Complementary material to the paper including tables, plots and identification times as well as additional studies over other databases can be found at http://sci2s.ugr.es/DPDDFF and in the associated Technical Report [28].

2. Preliminaries

A fingerprint is a pattern of valleys and ridges located on a fingertip. Although there are several ways to perform a matching between two fingerprints, many matching algorithms use the minutiae [3,7,29], comparing two minutiae sets to return a similarity score. The matching is performed once for each comparison between two fingerprints. Some of the existing matching algorithms offer very good matching precision [9], and others provide a fast response with slightly diminished accuracy [8], according to the taxonomy and results presented in [7].

There are two main variants of the fingerprint recognition problem [3]. Verification [4] is a 1:1 comparison to check if two fingerprints represent the same finger. Identification [5] consists of determining which fingerprint in a database of previously captured and stored templates $T = \{T_1, T_2, \ldots, T_n\}$ corresponds to a given input fingerprint I. An identification algorithm compares I to every T_i and returns the identity with the best matching score as shown in Eq. 1, where $Q(I, T_i)$ is the matching function. Thus, identification is a 1:n comparison:

$$\text{Identity} = \arg\max_i Q(I, T_i) \quad i \in \{1, 2, \ldots, n\} \tag{1}$$

This paper is focused on identification. Section 2.1 explains the current proposals for fast and scalable identification within large databases. Then, Section 2.2 presents the previous work about fingerprint fusion to improve the identification accuracy.

2.1. Scalable fingerprint recognition in large databases

The bottleneck of an AFIS when attempting to identify within a large database is the matching algorithm. Several proposals in the literature aim to overcome this problem.

FPGA-based systems implement the matching into a Field Programmable Gate Array [18,30], a hardware device that performs some operations very quickly, so that the overall identification time is reduced. Other approaches reduce the penetration rate in the database by using a previous classification or indexing step [31–34]. Nevertheless, in large databases this step may become the bottleneck, and the size of the subsets can become too large. Accuracy is degraded when the penetration rate is too small or the collision rate too high [33].

HPC is a common solution for reducing high execution times [19,20]. By using q computers with c cores each to perform a parallel search, the execution time can be reduced by up to a factor of qc. Moreover, the availability of more RAM memory allows more template fingerprints to be kept in a fast access device, avoiding slow access to secondary memory. Therefore, an adequate parallel framework can constitute a suitable tool for solving the identification problem in large databases [23–25].

2.2. Fingerprint information fusion

This section introduces two of the main trends to improve the accuracy of fingerprint recognition. On the one hand, the use of several fingers [13] increases the distinctiveness of the identities and tries to avoid the difficulties posed by injured fingertips or low quality scans. The matching function for f fingerprints becomes of the form $Q(\mathcal{I}, \mathcal{T}_i)$ where $\mathcal{I} = \{I_j \mid j \in \{1, \ldots, f\}\}$ and $\mathcal{T}_i = \{T_{ij} \mid j \in \{1, \ldots, f\}\}$. This approach has been successfully applied over latent fingerprints, which are of very low quality [35].

On the other hand, the combination of several matchers [14,36] aims to profit from their advantages, while leaving aside their weaknesses. Multi-algorithm techniques work in a similar way as multi-finger ones, so that the fused score obtained for f algorithms is $Q(I, T_i) = \mathcal{F}\big(Q_1(I, T_i), \ldots, Q_f(I, T_i)\big)$, where $\mathcal{F}$ is an aggregation function.

Multi-finger and multi-algorithm approaches can be categorized together according to the type of fusion they perform:

- **Feature fusion** [15,16,37,38]: this approach merges all f fingerprints of an identity into a single structure, which is compared