Knowledge-Based Systems 000 (2017) 1–13

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

# Distributed incremental fingerprint identification with reduced database penetration rate using a hierarchical classification based on feature fusion and selection

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

[a] Department of Internal Medicine, Ghent University, Ghent, Belgium
[b] Data Mining and Modelling for Biomedicine group, VIB Center for Inflammation Research, Ghent, Belgium
[c] School of Computer Science, University of Nottingham, Jubilee Campus, Wollaton Road, Nottingham NG8 1BB, United Kingdom
[d] Department of Applied Mathematics, Computer Science and Statistics, Ghent University, Ghent, Belgium
[e] Department of Computer Science and Artificial Intelligence of the University of Granada, 18071 Granada, Spain
[f] Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia

ARTICLE INFO

*Article history:*
Received 21 September 2016
Revised 16 March 2017
Accepted 18 March 2017
Available online xxx

*Keywords:*
Fingerprint recognition
Fingerprint identification
Fingerprint classification
Large databases
Feature selection
Hierarchical classification

ABSTRACT

Fingerprint recognition has been a hot research topic along the last few decades, with many applications and ever growing populations to identify. The need of flexible, fast identification systems is therefore patent in such situations. In this context, fingerprint classification is commonly used to improve the speed of the identification. This paper proposes a complete identification system with a hierarchical classification framework that fuses the information of multiple feature extractors. A feature selection is applied to improve the classification accuracy. Finally, the distributed identification is carried out with an incremental search, exploring the classes according to the probability order given by the classifier. A single parameter tunes the trade-off between identification time and accuracy. The proposal is evaluated over two NIST databases and a large synthetic database, yielding penetration rates close to the optimal values that can be reached with classification, leading to low identification times with small or no accuracy loss.



## 1. Introduction

Biometrics is a hot research field with many recent advances in different areas [1], such as signature [2], face [3] or fingerprint [4] recognition, template encryption [5], multimodal fusion [6] or spoofing detection [7]. They have been increasingly used during the last decades to replace tokens and passwords in many contexts. Fingerprint recognition is nowadays one of the most widespread methods for biometric identification [8]. It can be approached from two points of view: verification [9] aims to assess whether two fingerprints are taken from the same finger, whereas identification [10] consists of a search for a particular input fingerprint throughout a database of template fingerprints.

The scientific literature comprises some fast, accurate matching algorithms for verification [11–14]. However, the identification of an unknown input fingerprint among a large database of templates is still a challenging problem, both in terms of accuracy and runtime [15]. The number of runs of the matching algorithm increases linearly with the number of template fingerprints. This leads to an increase of the probability of erroneous identification, due to the higher number of non-matching templates that may cause false positives, and to a linear increase of the identification time, which eventually becomes unacceptably large.

High Performance Computing (HPC) [16] puts together massive computing resources (such as CPUs and GPUs) to reduce the time needed for computationally expensive tasks. It has been steadily applied to palliate the large identification time problem [4,17,18]. However, for sufficiently large databases, or when limited resources and budgets come into play, an approach from a pure HPC point of view may not be enough to obtain an efficient Automatic Fingerprint Identification System (AFIS).

A solution for such cases consists of reducing the fraction of the search space that is explored for the identification, called the database penetration rate [19]. Two main families of methods can be distinguished [20]: 1) classification approaches [21] that divide

* Corresponding author.
*E-mail addresses:* daniel.peralta@irc.vib-ugent.be (D. Peralta), Isaac.Triguero@nottingham.ac.uk (I. Triguero), salvagl@decsai.ugr.es (S. García), yvan.saeys@ugent.be (Y. Saeys), J.M.Benitez@decsai.ugr.es (J.M. Benitez), herrera@decsai.ugr.es (F. Herrera).

http://dx.doi.org/10.1016/j.knosys.2017.03.014
0950-7051/

Please cite this article as: D. Peralta et al., Distributed incremental fingerprint identification with reduced database penetration rate using a hierarchical classification based on feature fusion and selection, Knowledge-Based Systems (2017), http://dx.doi.org/10.1016/j.knosys.2017.03.014

the template database into a fixed number of classes, so that the input fingerprint is only searched within its own class, and 2) indexing methods [22] that map each fingerprint on a single or multi-dimensional space so that several impressions of the same fingerprint are mapped close to each other.

Classification is the most widespread of these two [21,23]. Most current proposals are based on the five classes defined by Henry [24], which are distinguished according to their ridge pattern and have an uneven distribution: *Arch, Left Loop, Right Loop, Tented Arch* and *Whorl*. The global features of the fingerprint (such as orientations maps or singular points) are extracted for the classification, and encoded as a numerical vector. Then, this vector is used within a machine learning algorithm to perform the classification itself.

Some of these feature extraction methods reject the fingerprints that do not comply with certain quality criteria [25–27]. This behavior is oriented towards avoiding a harmful reduction of the penetration rate that may impede finding the correct identity after a misclassification [20]. However, the rejected fingerprints cannot be classified, restraining the reduction of the penetration rate for the posterior identification.

Another common way of improving the classification consists of joining different types of features of the image [26,28–30]. Nevertheless, the processing of fused information also becomes more complex due to the larger amount of information that must be taken into account. This situation is typical for preprocessing methods [31], and more specifically feature selection algorithms [32,33], which aim to simplify the data by eliminating noisy or redundant information. Such techniques have already been successfully applied to the problem of fingerprint classification [34].

To overcome these problems, we propose a complete identification system that includes a hierarchical classification framework that fuses the information of multiple feature extractors. The classification accuracy is further improved by incorporating a feature selection step. With this method we aim to maximize the reduction of the penetration rate via classification, without rejecting fingerprints whose feature extraction may have been affected by low-quality impressions. The proposed AFIS is designed to efficiently and accurately perform identifications in large fingerprint databases. The process is split into two steps:

- First, a novel hierarchical classification framework combines different sets of features and classifiers, organized into layers, depending on which feature extractors reject a certain fingerprint, eliminating the rejection whilst still benefiting from the high accuracy provided by feature extractors with rejection. Furthermore, a feature selection process is applied to eliminate redundancies and maximize the accuracy within each layer. The final outcome of the hierarchical classifier for a given an input fingerprint is an ordering of the classes according to the probability of its membership.
- Then, a distributed identification is performed, which explores the different classes in the order given by the classifier to obtain an optimized trade-off between runtime and accuracy.

The AFIS described in this paper is evaluated on three different databases: NIST-SD4, NIST-SD14 and a large synthetic database. Different feature extractors and classifiers from the scientific literature were combined to measure their accuracy, along with that of the method proposed in the current work. The impact of the classification on the posterior identification process is also tested in terms of identification accuracy and runtime.

This paper is organized as follows. Section 2 presents previous work on fingerprint recognition and penetration rate reduction. Section 3 describes the proposed hierarchical classifier and identification procedure. Section 4 details the performed experiments and analyzes the obtained results. Finally, Section 5 offers the conclusions of the study.

## 2. Background

The fingerprint recognition problem is described in Section 2.1. Then, Section 2.2 carries out an overview of high performance computing for fingerprint recognition. Finally, Section 2.3 presents the approaches in the current literature to reduce the penetration rate of the search in fingerprint identification.

### 2.1. Fingerprint recognition

Some of the properties that make fingerprints suitable for recognition purposes have been known for more than a century [24]. However, it is in the last decades, with the development of Automatic Fingerprint Identification Systems, that the full potential of fingerprints is being exploited. They are essentially patterns of ridges and valleys, from which different kinds of features can be extracted to automate the recognition process [20]. Minutiae (bifurcations and ridge endings) are the most widely used features for identification, due to their capabilities to efficiently discern fingerprints [35].

As for the variants of the fingerprint recognition problem, identification is inherently more complex than verification, since it involves searching for an input fingerprint throughout a database of $n$ template fingerprints. This complexity increases along with the size of the database. As $n$ grows, the number of comparisons that must be performed increases linearly, and so does the identification time. The accuracy of the identification is also degraded due to the higher probability of false positives. The identification time and accuracy are tightly coupled objectives: very precise matching algorithms are usually computationally expensive, and faster approaches are less accurate [35]. Therefore, any AFIS needs to find a trade-off between accuracy and identification time.

### 2.2. High performance computing for fingerprint identification

HPC is a common tool to speed up computationally intensive tasks. Both multi-CPU and GPU systems have been employed to optimize the fingerprint identification process [4,17,18,36]. However, carrying out an effective and efficient application of HPC resources over such a problem is not straightforward. The database must be correctly distributed among the underlying hardware for a maximum speedup, and the parallel processes must be efficiently synchronized to minimize the overhead.

The use of a penetration reduction procedure increases the complexity of this task: each part of the database that can be potentially explored in parallel should be evenly distributed among the processors for a maximum performance [37]. Additionally, the processing of such parts should be performed so as to minimize the required synchronization.

### 2.3. Approaches for the reduction of the database penetration rate

Another way to speed up the identification time is by reducing the penetration rate of the search, which is the ratio between the number of fingerprints explored for the identification and the total size of the database [19]. There are two main approaches for this purpose: indexing (Section 2.3.1) and classification (Section 2.3.2).

#### 2.3.1. Indexing

Indexing methods can be based on various features, such as minutiae [22,38], orientations [39] or ridge frequencies [40], and usually do not consider rejecting fingerprints. They transform each fingerprint into a point in some multi-dimensional space, in such a way that several impressions of the same fingerprint ideally become very close points. Therefore, the input fingerprint is only compared with templates that fall close to that point. The search

Please cite this article as: D. Peralta et al., Distributed incremental fingerprint identification with reduced database penetration rate using a hierarchical classification based on feature fusion and selection, Knowledge-Based Systems (2017), http://dx.doi.org/10.1016/j.knosys.2017.03.014