Information Fusion 44 (2018) 57–64

Contents lists available at ScienceDirect

Information Fusion

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

Multiple classifiers in biometrics. part 1: Fundamentals and review

Julian Fierrez*, Aythami Morales, Ruben Vera-Rodriguez, David Camacho

School of Engineering, Universidad Autonoma de Madrid, Madrid, Spain

ARTICLE INFO

Keywords:
Classifier
Fusion
Biometrics
Multimodal
Adaptive
Context

ABSTRACT

We provide an introduction to Multiple Classifier Systems (MCS) including basic nomenclature and describing key elements: classifier dependencies, type of classifier outputs, aggregation procedures, architecture, and types of methods. This introduction complements other existing overviews of MCS, as here we also review the most prevalent theoretical framework for MCS and discuss theoretical developments related to MCS.

The introduction to MCS is then followed by a review of the application of MCS to the particular field of multimodal biometric person authentication in the last 25 years, as a prototypical area in which MCS has resulted in important achievements. This review includes general descriptions of successful MCS methods and architectures in order to facilitate the export of them to other information fusion problems.

Based on the theory and framework introduced here, in the companion paper we then develop in more technical detail recent trends and developments in MCS from multimodal biometrics that incorporate context information in an adaptive way. These new MCS architectures exploit input quality measures and pattern-specific particularities that move apart from general population statistics, resulting in robust multimodal biometric systems. Similarly as in the present paper, methods in the companion paper are introduced in a general way so they can be applied to other information fusion problems as well. Finally, also in the companion paper, we discuss open challenges in biometrics and the role of MCS to advance them.

1. Introduction

The basic aim of pattern recognition is to devise automatic procedures that maximize certain criteria for the recognition problem at hand, usually related to the recognition performance. This is normally achieved by comparing different existing pattern recognition algorithms on the specific problem studied, and selecting the best of them. Worth noting, by observing the errors misclassified by the different approaches, one can observe that some recognition errors committed by the best approach can be well resolved by the inferior methods. These observations motivated a big interest in combining classifiers in the 90's [1], which was followed by very active research since then. This is exemplified by the successful series of Workshops on Multiple Classifier Systems (MCS), conducted yearly since 2000 [2,3].

This multiple classifier approach can be found with different names in the literature [4]: classifier combination, classifier fusion, mixture of experts, committees of neural networks, consensus aggregation, expert conciliation, voting pool of classifiers, dynamic classifier selection, composite classifier design, classifier ensembles, divide-and-conquer classifiers, etc.

In addition to important theoretical advances, the above mentioned research in multiple classifier systems has resulted in highly successful practical developments in almost any field in which pattern classifiers are used, e.g., analytics of data streams [5], astronomy [6], biometrics person recognition [7], computer vision and medical image analysis [8], decision making [9], document analysis [10], hybrid systems [11], machine learning [12], neural information processing [13], and many others. One prototypical example of a big practical MCS success is the Viola-Jones cascade classifier [14], one of the most cited and widely used approaches in computer vision.

In the previous paragraph and related literature [4], the reader can find excellent surveys of MCS methods and algorithms. Out of those previous general references, the most related publications are the excellent MCS overview by Polikar [9], which is still a valuable reference after more than 10 years, and the quite recent overview of MCS applied to biometrics by Lumini and Nanni [15]. We complement the first overview by Polikar being more general, up to date, and more focused into fundamentals. On the other hand, in [9] one can find introductory descriptions of specific MCS algorithms like the ones only mentioned here in Table 1. With regard to the recent overview by Lumini and Nanni [15], here we are more comprehensive in our review of methods for biometrics, including the basics of the most prevalent theory of MCS applied to biometrics. We also develop recent trends and developments not discussed by Lumini and Nanni, including adaptive architectures

* Corresponding author.
E-mail address: julian.fierrez@uam.es (J. Fierrez).

https://doi.org/10.1016/j.inffus.2017.12.003
Received 24 November 2017; Received in revised form 19 December 2017; Accepted 21 December 2017
Available online 22 December 2017
1566-2535/

Table 1
Strategies in multiple classifier systems. Adapted from Maltoni et al. [34].

Method	Architecture	Level	Train.	Adapt.	Comments
Class set reduction	Serial/Parallel	Rank/Conf.	Yes	No	Efficient
Voting, AND/OR	Parallel	Abstract	No	No	Assumes independency
Associative switch	Parallel	Abstract	Yes	Yes	Explores local expertise
Borda count	Parallel	Rank	Yes	No	Converts ranks to confidences
Logistic regression	Parallel	Rank/Conf.	Yes	No	Converts ranks to confidences
Dempster–Shafer	Parallel	Rank/Conf.	Yes	No	Fuses non-probabilistic scores
Prod, min, max	Parallel	Confidence	No	No	Assumes independency
Sum, median	Parallel	Confidence	No	No	Assumes independency; robust
Gen. Ensemble	Parallel	Confidence	Yes	No	Considers error correlations
Stacking	Parallel	Confidence	Yes	No	Exploits scarcity in data
Fuzzy Integrals	Parallel	Confidence	Yes	No	Fuses non-probabilistic scores
Bagging	Parallel	Confidence	Yes	No	Needs many classifiers
Random subspace	Parallel	Confidence	Yes	No	Needs many classifiers
Adaptive weighting	Parallel	Confidence	Yes	Yes	Explores local expertise
MLE	Parallel	Confidence	Yes	Yes	Explores local expertise
Boosting	Parallel/Hier.	Abstract	Yes	No	Needs many classifiers
Neural tree	Hierarchical	Confidence	Yes	No	Handles many classes
Hierarchical MLE	Hierarchical	Confidence	Yes	Yes	Explores local expertise

and practical algorithms implementing the discussed new trends.

In order to be as self-contained as possible while avoiding overlap with related publications, this paper is divided into two Parts, each of them with slightly different intended audience.

In the present paper, Part 1, we first provide a brief introduction to MCS outlining basic nomenclature, architecture, and key elements, with a focus into the fundamentals of MCS. We refer the reader to the references in previous paragraphs for descriptions of established MCS methods and algorithms.

After the brief introduction to MCS, we also review here in Part 1 the application of MCS to the particular field of multimodal biometric person authentication in the last 25 years or so, as a prototypical area in which MCS has resulted in important achievements. We review MCS in multimodal biometrics with general descriptions of main MCS elements, methods, and algorithms; facilitating the export of experiences and methods to other information fusion problems.

The companion paper, Part 2 of this series of two papers, is intended for researchers knowledgeable in MCS interested in recent developments in context-based information fusion [16] coming from the biometrics research community, or newcomers to MCS that have first addressed Part 1.

The companion paper ends with a discussion of open challenges in biometrics that can be addressed and advanced using MCS. The challenges exposed largely follow the excellent survey and outlook of the field of biometric person recognition by Jain et al. [17], which we complement with our personal view, and augment with the way MCS developments can advance key challenges in biometrics.

Biometrics person recognition shares many issues and challenges with other pattern recognition applications like video surveillance [18], speech technologies [19], human-computer interaction [20], data analytics applications [21], behavioral modeling [22], or recommender systems [23]. By keeping our discussion on MCS methods in biometrics as general as possible, we hope to provide some hints for potential research and advancements in other pattern recognition and information fusion areas as well.

The present paper is organized as follows. Section 2 provides an introduction to Multiple Classifier Systems (MCS), including: nomenclature, architecture, summary of classical techniques, and a flexible theoretical framework. Section 3 overviews the application of MCS techniques to biometrics in the last 25 years, with emphasis on the most popular architecture, namely: post-classification fusion; which is further discussed separately for combination and classification approaches. We complete Section 3 with a discussion of score normalization, and its application to MCS. The paper ends in Section 4 with some concluding remarks.

2. Multiple classifier systems (MCS)

Multiple classifier approaches can be categorized depending on: assumption about classifier dependencies, type of classifier outputs, aggregation procedure, and architecture.

Classifier dependencies. In general, we may have different classifier outputs because of [24]: different feature sets, different training sets, different classification methods, different parameters in the classification method, or different training sessions. All these reasons result in a set of classifiers whose outputs may be combined with the hope of improving the overall classification accuracy. Classifier combination is specially useful if the individual classifiers are largely diverse [25]. If this has not been guaranteed by the use of different training sets, resampling techniques like rotation or bootstrap may be used to artificially create such differences. Examples of classifier combination based on resampling strategies are the well known stacking [26], bagging [27], and boosting [28].

In the case of multimodal biometric authentication, the independence between classifiers (one for each modality) is normally assumed.

Type of classifier outputs. The outputs of the different classifiers can be classified into three levels [29]: 1) abstract, 2) rank, and 3) measurement (or confidence). At abstract level, each classifier only outputs a class label. At rank level, each classifier outputs a ranked list of classes, with the class ranked first being the first choice. At measurement level, each classifier outputs a numerical value indicating the belief or probability that the pattern belongs to a given class.

Aggregation procedures. Aggregation procedures can be first classified according to trainability and adaptivity. Some combiners do not require training while others are trainable. The trained combiners may lead to better performance at the cost of additional training data and additional training. Some combiners are adaptive in the sense of weighting the contribution of each expert depending on the input pattern. Conversely, nonadaptive combiners consider all input patterns in the same way. Adaptive schemes can exploit the detailed error characteristics of the individual classifiers under different input patterns. Examples of adaptive combination strategies include adaptive weighting [30], mixture of local experts (MLE) [31], and hierarchical MLE [32].