N°: 103/2017_C/ELC

PEOPLE'S DEMOCRATIC REPUBLIC OF ALGERIA
Ministry of Higher Education and Scientific Research

UNIVERSITY OF SCIENCE AND TECHNOLOGY HOUARI BOUMEDIENE

FACULTY OF ELECTRONICS AND COMPUTER SCIENCE

THESIS

Presented for the requirement of the degree of

Doctor

IN : ELECTRONICS

Speciality : **Telecommunications and Information Processing**

By : **Bilal HADJADJI**

Contribution of Using One-class Classification Combination for Multi-class Pattern Recognition

Mr	LARABI	Slimane	Prof.	USTHB	President
Mr	CHIBANI	Youcef	Prof.	USTHB	Thesis Supervisor
Mrs	OUaRZEDDINE	Mounira	Prof.	USTHB	Examiner
Mr	BELHOCINE	Mahmoud	Dir. R	CDTA	Examiner
Mr	MEKAOUI	Slimane	Prof.	USTHB	Examiner
Mr	DAAMOUCHE	Abdelhamid	MCA	Univ. Boumerdès	Examiner

To all who helped pried and wished for me good luck

Acknowledgments

The work presented in this thesis is carried out at the Laboratoire d'Ingénierie des Systèmes Intelligents et Communicants (LISIC) Faculty of Electronics and Computer Science of University of Science and Technology Houari Boumediene (USTHB). This thesis aims at the preparation of a PhD in Telecommunications and Information Processing.

Foremost, I would like to express my sincere gratitude to my advisor Prof. Youcef Chibani for the continuous support of my Ph.D study and research, for his patience, motivation, immense knowledge and notably enthusiasm. His guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better supervisor and mentor for my Ph.D study.

I wish to express my gratitude to Mr. LARABI Slimane, Professor at USTHB who will chair the jury of my thesis, which is a great honor for me.

My warm thanks go to the members of the jury who have agreed to spare their time to examine my work: Mrs. OUERZEDDINE Mounira, Professor at USTHB, Mr BELHOCINE Mahmoud, Research director at CDTA, Mr MEKAOUI Slimane, Professor at USTHB and DAAMOUCHE Abdelhamid MCA at University of Boumerdès.

My sincere thanks also go to all members of the LISIC laboratory, especially Abdenour Sehad and Hassiba Nemmour for their help and support.

I also want to thank my friend Lotfi Achiri, Khetab Brahim, Korbab Mohamed, Ali Belahcen , Amar Tirchi and Amar Lacet.

I would like to express my sincere thanks and warm gratitude to my family for their support and encouragement that have allowed me to reach the end of this thesis. I hope they will be proud of me for this work.

And at last but not the least, I would like to thank my future wife Hadjer Lakhder-ezzine and all her family.

To all the above-mentioned individuals and all the others that I have not mentioned but who I know, from my heart, thank you!

Abstract

Usual multi-class classifiers take in consideration either the entire classes for generating the classification model, such as the neural networks, or by splitting up the original problem into a set of two-class sub-problems. In contrast, the One-Class Classifier (OCC) takes into account only the target class, which allows easily achieving the multi-class implementation. However, although these important flexibility of using the OCC for solving multi-class classification problem, handling a single system of OCC usually achieves less accuracy than the usual multi-class implementations. Thus, this thesis focuses for proposing to explore multiple classifier systems based on OCC leading to perform higher accuracy and keep at the same time the open classification system. Also, thesis contributions can be summarized as follows:

- A new Dynamic Weighted Average (DWA) combination rule is proposed for combining different types of OCC for multi-class classification. Experimental results conducted on several real-world datasets prove the effective use of the proposed approach where the DWA rule achieves the best results against fixed rules as well as the decision template. Furthermore, comparison of the proposed open classification system based on the K-Nearest Neighbor classifier shows the superiority of the proposed open multi-class classifier.
- An improved one class Auto Associative Neural Network (AANN) ensemble is proposed based on a selection algorithm for selecting the appropriate training samples. The proposed framework allows enhancing the AANN combination and classification robustness. Experiments conducted on several real-world datasets prove the efficiency of the proposed algorithm for more robust AANN ensembles.
- A new combination scheme of OCCs is proposed based on Fuzzy Integral (FI) operators. Furthermore, an alternative framework is proposed to design a parameter-independent and open-lexicon handwritten Arabic word recognition system as well as a new density measure function. Experimental results conducted on Arabic handwritten dataset using different types of OCCs with large number of classes show the superiority of FI for OCC ensembles.
- An Open Handwritten Signature Identification System (OHSIS) is proposed by using conjointly the Curvelet Transform (CT) and the One-Class classifier based on Principal Component Analysis (OC-PCA). A combination based on Choquet fuzzy integral is explored to combine multiple individual OHSISs. Experimental results conducted on standard CEDAR and GPDS handwritten signature datasets report 97.99% and 94.96% correct identification rate, respectively, which outperform the state-of-the-art when using few reference signatures.
- Two stage combination system is proposed for writer identification based on handwriting fragments. The first stage is devoted for fragment combination based on a proposed Dynamic Fragment Weighting Combination (DFWC) rule. On the other hand, the second stage is dedicated for combining different writer identification systems fed by three descriptors via the FI combination strategy. Experimental results conducted on the well-known IFN/ENIT and IAM databases show good adaptation of the OCC with DFWC. Moreover, the Choquet combination scheme offers more improvements to achieve 97.56% and 94.20% for the used databases, respectively. The obtained results highlight the reliability of the proposed system in comparison with recent studies for writer identification issue.

Table of Content

Chapter 2

Chapter 6

List of Tables

List of Figures

List of Symbols

x	Test pattern
μ	Mean of the training samples
Σ	Covariance matrix calculated from the training samples.
$th^{Gaussian}$	Threshold for one-class Gaussian classifier
$f_j^{Gaussian}$	Individual Gaussians
φ_j	Mixing coefficients of mixture of
N_{MoG}	Number of Gaussians
x_i	Individual training objects
Σ_i	Diagonal covariance matrices
th^{Parzen}	Threshold for one-class Parzen classifier
ε_{kc}	Error function for K-Means classifier
u_j	Centers of k-Centers classifier
S	Training set.
f^{NN}	Density function of nearest neighbor classifier
V	Smallest volume of the spheres
$f^{SVM}(x)$	Decision function of OC-SVM
$K(.,.)$	OC-SVM kernel
α_i	Lagrange multipliers
ν	Percentage of samples considered as outliers
γ	Kernel parameter
u_j	Prototype vectors

$\mathrm{th}^{\mathrm{K-Means}}$	Threshold for one-class K-Means classifier
K	number of clusters of K-means classifier
K	number of centers of k-center classifier
p	Eigenvectors
x_{proj}	Projection of the test pattern x
n	Number of training samples
th^{AANN}	Threshold value of AANN classifier
$Er(x)$	Reconstruction error
C	Set of classes
m	Number of classes
L	Number of classifier
$P_i\ (c_j\,/x)$	A posteriori probabilities
λ	Fuzzy measure parameter
z_i	Information source
g^i	Density measure
$h(z_i)$	Objective evidence
$g(A_i)$	Fuzzy measure
α	OR operator parameter
β	AND operator parameter
I_S	Sugeno fuzzy integral
I_C	Choquet fuzzy integral
I_{S-AND}	Sugeno-AND fuzzy integral

I_{C-AND}	Choquet-AND fuzzy integral
I_{OR}	OR fuzzy integral
$f_j^{OCC\ i}$	OCC model of i^{th} classifier and j^{th} class
DT_j	Decision template
DP	Decision Profile
d_E	Euclidian distance
w_i	Weight associated to the i^{th}classifier
r_i	Average error rate
$y(x)$	Label of a test sample x
$w_i(x)$	Weight assigned to the i^{th} classifier for a test sample x
$op_i^j(x)$	j^{th} class of i^{th} classifier response
op_i^{max}	i^{th} classifier maximum response
S_{pr}	Pertinent training sample set
S_{out}	Outlier set
f_{ent}	Representative model of AANN trained on the entire training set
S_{ord}	Ordered training samples
d_{ij}	OCC output value
dm_{ij}	Average of training samples values
δ	Calibrating parameter
$C_{j,k}$	Curvelet coefficients computed at the scale j and the orientation k.
E	Energy of curvelet coefficients
$N \times N$	Size of small writing fragments

P	Run-length matrix
ϕ	Quantized orientations
C_{ij}	Filter response value
Q	Query document
q_i	Test fragment
$\bar{f}_m^{OCC}$	Average of all reference fragments outputs
r_i	Reference fragment
τ	Calibrating parameter
$W_m(q_i)$	Appropriate weight dynamically for each test fragment q_i

List of Acronyms

Bi-SVM	Binary Support Vector Machine
OAO	One-Against-One
OCC	One-Class Classifier
HMM	Hidden Markov Models
MoG	Mixture of Gaussians
OC-NN	One-Class Nearest Neighbor
OC-SVM	One-Class Support Vectors Machine
PCA	Principal Component Analysis
AANN	Auto-Associative Neural Networks
SVDD	Support Vector Data Description
K-NN	K-Nearest Neighbor
SOM	Self Organization Map
PCAD	PCA Description
LPD	Linear Programming Description
IDS	Intrusion Detection Systems.
FKNN	Fuzzy K-Nearest Neighbor
KPCA	Kernel Principle Component Analysis
ECOC	Error Correcting Output Code
DT	Decision Template
FI	Fuzzy Integral
OWA	Ordered Weighted Averaging

SWA	Static Weighted Average
DWA	Dynamic Weighted Average
MCS	Multiple Classifier System
AERs	Average Error Rates
CEDAR	Center of Excellence for Document Analysis and Recognition
STS-AANN	Selected Training Samples Auto Associative Neural Networks ()
MRR	Mean Recognition Rate
DSmT	Dezert-Smarandache Theory
CT	Curvelet transform
Avg	Average
Prod	Product
Min	Minimum
Max	Maximum
OHSIS	Open Handwritten Signature Identification System
C-FI	Choquet Fuzzy Integral
GPDS	Grupo de Procesado Digital de Senales
IR	Identification Rate
ES	Equi-Spaced
EM	Equi-Mass
DFWC	Dynamic Fragment Weighting Combination
LBP	Local Binary Pattern
LPQ	Local Phase Quantization
LTP	Local Ternary Patterns

oBIF	oriented Basic Image Feature
RL	Run length
BIF	Basic Image Feature
STFT	Short Term Fourier Transform

Introduction

Pattern recognition is a branch of machine learning that focuses on the recognition of patterns and regularities in data. In particular, pattern classification is related to various real-world applications such as the classification of a given email to "spam" or "non-spam". Moreover, the classification problem becomes more complex when the number of classes is increased such as in the application of person identification based on biometric information (Schwartz *et al*., 2012). The classifier bloc allows to convert an input feature vector into class vector and therefore attribute it to one of a given classes.

Generally, a pattern recognition system is composed of four main modules: acquisition, preprocessing, feature generation and classification. This later allows assigning a pattern to predefined class. Various classification methods have been proposed as template matching techniques (Deng et al., 1999; Fang et al., 2003), minimum distance classifiers (Fang et al., 2001; Sabourin et al., 1997), support vector machine (Justino et al., 2005), hidden Markov models (Justino et al., 2001; Justino et al., 2005), neural networks (Kaewkongka et al., 1999 ; Quek et al., 2002), etc. However, although the good offered classification accuracy, these techniques perform closed system. Indeed, usual multi-class classifiers (Duda et al., 2001) take in consideration either the entire classes for generating the classification model, such as the neural networks, or by splitting up the original problem into a set of two-class sub-problems. For instance, the Binary Support Vector Machine (Bi-SVM) uses discriminant functions to make a multi-class decision out of binary decisions. Hence, various implementations can be performed as the One-Against-One (OAO) and One-Against-All (OAA) and directed acyclic graph (Hsu and Lin, 2003). In contrast, the One-Class Classifier (OCC) takes into account only the target class, which allows easily achieving the multi-class implementation. Indeed, the simple use of the OCC for multi-class implementation has been investigated in many applications such as sound classification (Rabaoui et al., 2009), cognitive state classification of brain activity (Boehm et al., 2011) and so on. Thus, using OCC to achieve a multi-class implementation allows serving to resolve two main problems. Firstly, the multi-class implementation based on OCCs is considered as open schme since each classifier is trained only on its target class offering thus an extensible scheme. Indeed, adding a new class to the multi-class scheme does not require retraining it again on all classes

It requires training only the added classifier. Secondly, this implementation offers less computational cost in terms of training time and memory space against other implementations such as OAO and OAA based on the Bi-SVM classifier (Yeh et al., 2009).

In the last decades, OCC has attracted more attention for many researchers leading to use it for solving the multi-class classification problem (Goh et al, 2005; Ban et al,, 2006; Yeh et al., 2009; Rabaoui et al,, 2009; Boehm et al,, 2011 ; Seokho et al,, 2015). Those are motivated by the advantages offered by the OCC against the usual implementations based on bi-class or multi-class classifiers (Yeh et al., 2009). However, using a single system of OCC for the multi-class implementation usually achieves less accuracy than the usual multi-class implementations.

This drawback is due to the absence of information about the other classes which causes overlap between the different models. Thus, the OCC will estimate a large boundary that encloses areas of the feature space where the positive class has low density, resulting in much classification confusion (Shieh and Kamm, 2009). This can be highly problematic for many applications where the classification accuracy must be kept to a maximum.

Efficient way to produce an improved classifier is an ensemble method that combines multiple weak classifiers (Breiman, 1996). In this context, various ensemble methods can be performed, such as combining different one class classifiers (Tax and Duin, 2001) and training on different sets of features (Perdisci and Lee, 2006). However, OCC ensembles still await proper attention to provide better improved ensembles and also to be explored for more applications. Hence, a distinction should be highlighted between OCC ensemble for solving one-class problems, multi-class implementation and ensemble of multi-class implementation, which represents the hybrid approach.

Thus, this thesis proposes to explore multiple multi-class implementation systems based on OCC leading to perform higher accuracy and keep at the same time the open classification system. More precisely, the objective of the presented thesis is the enhancement of OCC ensembles for solving the multi-class classification problem. The proposed contributions are devoted for real world applications in general as well as specifically for some handwritten recognition applications such as word recognition, signature identification and writer identification.

Therefore, our contributions to the scientific advancement of OCC combination for multi-class pattern classification can be summarized as follows:

1. **Updating the state of art and proposing a new taxonomy of OCC combination:**

A taxonomy of OCC ensembles is presented where we differentiate between OCC ensemble for solving one class problems, multi-class implementation and ensemble of multi-class implementation defining the hybrid approach.

2. **Proposing architecture for combining different types of OCC, as well as new combination rule:**

Using OCCs for the multi-class implementation usually achieves less accuracy than the usual multi-class implementations. Hence, in order to improve the accuracy and keep the offered advantage, the suitable approach consists to combine different classifiers. Thus, in this thesis a study of combining different types of OCC for multi-class classification is proposed by means of a new dynamic weighted average combination rule.

3. **Improving the combination via reducing the effect of outliers applied to one-class based neural network classifier:**

The enhancement of classifier ensembles is always performed on the combination rule. In this thesis, we propose an alternative method for OCC ensembles improvement. The proposed approach relies on an effective selection of training samples for performing an improved Auto-Associative Neural Networks (AANN) ensemble. More precisely, the pertinent samples are selected according to their reconstruction error calculated between the input and its corresponding output generated by an initial model.

4. **Proposing the use of fuzzy integral operators for combining a same type of classifiers trained by different descriptors**

Fuzzy integral has been reported to give excellent results in comparison with fixed rules for classifiers combination. Thus, in this thesis, a combination scheme of OCCs is proposed based on fuzzy integral operators. Furthermore, the concept of the already proposed dynamic weighting is incorporated for proposing new and more adapted density measure function. Besides, an alternative framework is proposed to design a parameter-independent and open system. This scheme has been performed to design an open-lexicon handwritten Arabic word recognition system as well as an open offline handwritten signature identification system.

5. Proposing two combination stage scheme of clustered OCC for the writer identification

Two combination stage system of OCC is proposed for writer identification based on handwriting fragments. The first stage is devoted for fragment combination based on a proposed dynamic fragment weighting combination rule. On the other hand, the second stage is dedicated for combining different writer identification systems fed by three descriptors via the FI combination strategy.

The remaining of this thesis is organized in six (6) chapters as follows. Chapter 1 is devoted for presenting an overview of OCCs as well as classifier combination approaches. Chapter 2 proposes a combination of different types of OCCs as well as new combination rule for more OCC ensemble enhancement. Chapter 3 presents improved one-class neural network classifier ensembles by optimized selection of training samples. Chapter 4 explores the fuzzy integral for enhancing OCC ensembles which is used for open-lexicon Arabic word recognition system. Chapter 5 proposes an offline handwritten signature identification system using the Curvelet transform and ensemble of one-class principal component analysis. Chapter 6 is devoted for presenting two combination stages of OCCs for writer identification based on handwriting fragments. Finally, the conclusion and future works are provided in the last of this thesis.

Chapter 1

Review of One-Class Classifiers and Combination Approaches

Abstract

This chapter aims to present an overview of the most used OCC methods as well as the common combination approaches, notably the investigated methods in this thesis. Firstly, we recall the three based types of OCC used in pattern recognition namely density-based, boundary-based and reconstruction-based. After that, we browse the combination schemes used for designing multiple classifier systems, which differ mainly by the adopted architecture, the classifier combination level and the used combination rule. Moreover, we present a taxonomy of OCC ensembles where we distinguish between OCC ensemble for solving one class problems, multi-class implementation and ensemble of multi-class implementation, which represents the hybrid approach. In particular, the parallel classifier combination is studied, where an overview of the main combination rules based on fixed rules and fuzzy integral is given

1.1. Introduction

In pattern classification systems the classifier is a crucial bloc to perform an efficient classification. Thus, various types of classifiers have been proposed and explored for different recognition applications such as neural networks, Hidden Markov Model (HMM), Support Vector Machines (SVM), distances and so on. However, although the good offered the classification accuracy these techniques perform closed systems**.** Indeed, the One-Class Classifier (OCC) has the advantage to take into account only the target class, independently from other classes which allows achieving an open multi-class implementation. Hence, adding a new class to the multi-class scheme does not require retraining it again on all classes. It requires training only the added classifiers. However, using single system of OCC for the multi-class implementation usually achieves less accuracy than the usual multi-class implementations for instance when using the Bi-SVM. This is due to the absence of information about the other classes which causes overlap between the different models. Thus, the OCC will estimate a large boundary that encloses areas of the feature space where the positive class has low density, resulting in much classification confusion (Tax and Duin, 2001).

Ensemble method is an efficient way to produce an improved classifier that relies on combining multiple weak systems (Breiman, 1996). Various ensemble architecture and methods have been developed such as serial parallel and hybrid. Moreover, various combination rules can be explored according to the combination level.

Thus, this chapter presents an overview of the most used OCC methods as well as the common combination approaches, notably the investigated methods in this thesis. Furthermore, a taxonomy of OCC ensembles methods is proposed in this chapter.

The remaining of this chapter is organized as follows. Section 2 presents OCC definitions and mathematical background. In section 3, we present the classifier combination approaches. In section 4, we propose a taxonomy and review of OCC ensemble categories where we make difference between OCC ensemble for solving one class problems, multi-class implementation and ensemble of multi-class implementation, which represents the hybrid approach. Section 5 highlights explored fuzzy integral combination rules which is explored in the present thesis.

1.2. OCC definitions and methods

One Class classifiers are defined as a machine learning to model a restricted domain in a multi-dimensional pattern space using only a set of the target class (Tax, 2001). Thus, the OCC is considered one among the nearest approach to the human learning for the classification task, due to its ability to learn the model of each class independently of the remaining classes. Other synonymous terms used in the literature for the OCC include: outlier detection, novelty detection, concept learning and data description (Tax, 2001). The OCCs can be categorized into three types: (i) density-based; (ii) boundary-based and (iii) reconstruction-based method.

1.2.1. Density methods

The most straightforward method to obtain a one-class classifier is to estimate the density of the training data (Tarassenko et al., 1995) and to set a threshold on this density. Several distributions can be assumed, such as a Gaussian or a Poisson distribution, and numerous tests, are then available to test new objects (Barnett and Lewis, 1978). In this thesis, we will consider three density models, the normal model, the mixture of Gaussians and the Parzen density.

1.2.1.1. Gaussian model

This method assumes that the data is distributed according to the normal (Gaussian) distribution (Bishop, 1995). The mean and covariance matrix are estimated from the training data, and instances located in the two tails are considered outliers (Ullman, 1978), as shown in figure 1.1. The probability distribution for a d-dimensional object x is given by:

$$f^{Gaussian}(x) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right) - th^{Gaussian} \qquad (1.1)$$

where μ is the mean and $\sum$ is the covariance matrix calculated from the training samples. The method is very simple and it imposes a strict unimodal and convex density model on the data.

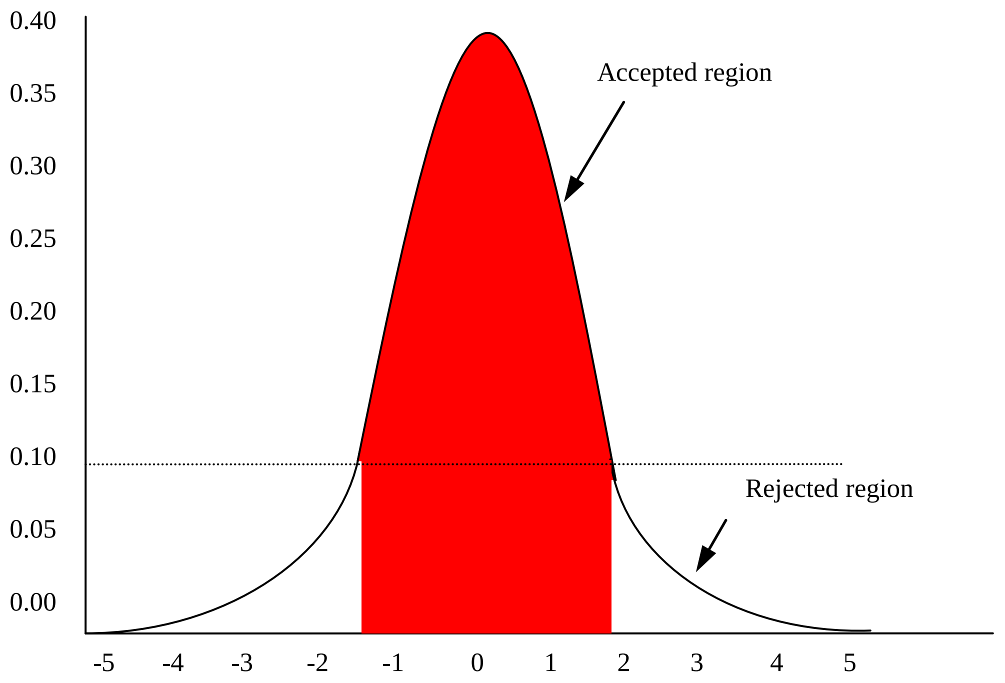


Figure 1. 1: Threshold on a 1-dimensional Gaussian distribution

1.2.1.2. Mixture of Gaussians

The Gaussian distribution assumes a very strong model of the data. It should be unimodal and convex. For most datasets, these assumptions are violated. To obtain a more flexible density method, the normal distribution can be extended to a Mixture of Gaussians (MoG) (Duda and Hart, 1973). A mixture of Gaussians is a linear combination of normal distributions (Bishop, 1995):

$$f^{MoG}(x) = \frac{1}{N_{MoG}} \sum_j \varphi_j f_j^{Gaussian}(x) - th^{MoG} \tag{1.2}$$

where φ_j is the mixing coefficient of individual Gaussians $f_j^{Gaussian}$ and N_{MoG} is number of Gaussians defined beforehand by the user. It has a smaller bias than the single Gaussian distribution, but it requires far more data for training, and thus shows more variance when a limited amount of data is available.

1.2.1.3. Parzen density estimation

The Parzen density estimation (Parzen, 1962), is also an extension of the previous method. Where the estimated density is a mixture of, most often, Gaussian kernels centered on the individual training objects x_i, with (often) diagonal covariance matrices $\sum_i = PI$. Thus, the classification can be performed as follows:

$$f^{Parzen}(x) = \frac{1}{N} \sum_i \frac{1}{(2\pi)^{d/2} |\sum_i|^{1/2}} \exp\left(-\frac{1}{2}(x - x_i)^T \textstyle\sum_i^{-1} (x - x_i)\right) - th^{Parzen} \tag{1.3}$$

The equal width P in each feature direction means that the Parzen density estimator assumes equally weighted features and it will, therefore, be sensitive to the scaling of the feature values of the data, especially for lower sample sizes. Hence, training a Parzen density consists of the determination of the optimal width of the kernel P.

1.2.2. Boundary methods

The boundary methods heavily rely on the distances between objects. They tend to be sensitive to scaling of the features. On the other hand, the number of required objects is smaller than in case of the density methods. Therefore, in this thesis, we consider the k-centers method, the One-Class Nearest Neighbor (OC-NN) and the One-Class Support Vectors Machine (OC-SVM).

1.2.2.1. K-centers

k-Center is a boundary method that relies on covering the training dataset with k smallest hyperspheres of equal radii (Ridler and Calvard, 1978). The centers of hyperspheres are placed on some observation vectors selected from the samples so that the following function would be minimized (Tax, 2001):

$$\varepsilon_{kc} = \max_{i=1}^{n}\left(\min_{j=1}^{k} \left\|x - u_j\right\|^2\right) \tag{1.4}$$

The error function ε_{kc} can be minimized using the batch algorithm allowing a best estimation of observation vectors. During classification, the distance between a new observation vector and centers u_j is calculated as:

$$f^{k-Center}(x) = -\min_{j=1}^{k} \left\|x - u_j\right\|^2 + th^{k-Center} \tag{1.5}$$

Therefore, the test sample x may either be rejected or accepted according to the threshold value $th^{k-Center}$ defined during the training or validation step.

1.2.2.2. Nearest Neighbor

According to (Tax and Duin, 2005), OC-NN relies on finding the distance of a test pattern x to its nearest neighbor in the training set S. The method attempts to estimate the density function as:

$$f^{NN}(x) = \frac{1/n}{V(\|x - NN^S(x)\|)} \tag{1.6}$$

Where n is the number of training samples and V defines the smallest volume value with the centre in x surrounding the observation vector nearest to x.

During classification, a test sample x is accepted when its local density is larger or equal to the local density of its first nearest neighbor in the training set. Otherwise, it is rejected.

1.2.2.3. Support Vector Machines

The OC-SVM introduced by Schölkopf algorithm relies on mapping the training samples belonging to S into a high dimensional feature space leading to find the maximal margin hyperplane, which best separates the training samples from the origin. For a test pattern x, a decision function namely $f^{SVM}(x)$ is performed for evaluating in which side of the hyperplane it falls. Formally, the decision function takes the following form:

$$f^{SVM}(x) = sgn\{\textstyle\sum_{i=1}^{n} \alpha_i K(x, x_i) - \rho\} \tag{1.7}$$

$K(.,.)$ defines the OC-SVM kernel that allows projecting data from the original space to the feature space. n is the number of training sample and α_i are the Lagrange multipliers computed by optimizing the following equations:

$$min_{\alpha} \left\{ \frac{1}{2} \alpha_i \alpha_j \ K(x_i, x_j) \right\} \tag{1.8}$$

Subject to,

$$0 \leq \alpha_i \leq \frac{1}{\nu n} \tag{1.9}$$

$$\sum_i^n \alpha_i = 1 \tag{1.10}$$

ν is the percentage of samples considered as outliers taking values in the range [0, 1]. A pattern x is then accepted when $f(x) > 0$. Otherwise, it is rejected. Various kernel functions can be used as polynomial or Radial Basis Function (RBF) or multilayer perceptron (Ridler and Calvard, 1978). Usually, the RBF kernel is the most used defined as:

$$K(x, x_i) = exp\big(-\gamma \ d(x, x_i)\big) \tag{1.11}$$

Such that $d(x, x_i)$ represents the distance between the test pattern and the training samples, which is defined as follows:

$$d(x, x_i) = \|x - x_i\|^2 \tag{1.12}$$

γ is the kernel parameter that allows controlling the distribution of the training samples in the feature space aiming to better separation.

1.2.3. Reconstruction methods

The reconstruction methods have been mainly constructed to model the data and not for one-class classification. The data is modeled by using prior knowledge and making assumptions about the data generating process, a model is selected and fitted to the data. New objects can then be described in terms of a state of the generating model. While using reconstruction methods, we assume that a more compact representation of the target data can be obtained and that, in this encoded data, the noise contribution is decreased. This representation then simplifies further processing without harming the information content (Tax, 2001).

In this thesis we will consider the following examples of reconstruction methods: the K-Means clustering, (Principal Component Analysis) PCA and (Auto Associative Neural

Networks) AANN. The methods differ in the definition of the prototypes or subspaces, their reconstruction error and their optimization routine.

1.2.3.1. K-Means

K-Means clustering is one of the simplest reconstruction methods. The main idea behind performing K-Means clustering for OCC is to assume that the data is clustered and can be described by a set of prototype vectors U_j (Tax and Duin, 2004). The number K of prototype vectors should be selected beforehand. In this thesis, the optimal K value is deduced from the evaluation of training samples. During classification, the decision function based on the reconstruction error is calculated as:

$$f^{K-Means}(x) = -\min_{j=1}^{K} \left\| x - U_j \right\|^2 + th^{K-Means} \tag{1.13}$$

Therefore, the test sample x may either be rejected or accepted according to the threshold value $th^{K-Means}$ defined during the training or validation steps. The placement of prototype vectors is derived from the training dataset, which are selected during the clustering for minimization purposes:

$$\varepsilon_{km} = \sum_{i \in S_j} \left(\min_{j=1}^{K} \left\| x - U_j \right\|^2 \right) \tag{1.14}$$

Both batch and on-line algorithms can be employed for finding the solution minimizing ε_{km}. Here, we use the batch algorithm, which consists to group at each step the observation vectors into K disjoint sets S_j according to the nearest prototype vectors. Then, the prototype vectors are recalculated as:

$$u_j = \frac{1}{N_j} \sum_{i \in S_j} x_i \tag{1.15}$$

N_j defines the cardinal of subset S_j and x_i is the training sample contained into S_j. This procedure is repeated until convergence.

This method requires defining the number of clusters K to be fixed beforehand. Therefore, K-Means clustering is sensitive to the correctness of the parameter K, if it differs from the real number of clusters in the data, then wrong clusters may be produced (Dietterich, 2000).

The K-Means clustering has been explored for anomaly intrusion detection (Marchette, 1999) for clustering networked computers into "activity groups", and also for reducing the raw data to a manageable size (Zanero and Savaresi, 2004).

1.2.3.2. Principal Component Analysis

Principal Component Analysis (PCA) is a statistical procedure that uses orthogonal transformation to convert a set of observations from possible correlated variables into a set of linear uncorrelated variables, called principal components. The PCA is often used to reduce the dimension of data since the number of principal components is less than or equal to the number of original variables. The simplest optimization procedure uses eigenvalue decomposition to compute the eigenvectors of the target covariance matrix. The eigenvectors with the largest eigenvalues define the principal axis (Tax, 2001).

In order to implement one-class classification, the PCA method relies on describing the target data by a linear subspace, which is represented by the eigenvectors of the data covariance matrix. Only p eigenvectors are stored in a matrix W of $d \times p$ (where d is the dimensionality of the original feature space). To check if a new object fits the target subspace, a decision function based on the reconstruction error is computed according to the difference between the original object x and its projection onto the subspace as follows (Tax, 2001):

$$f^{PCA}(x) = -\left\| x - x_{proj} \right\|^2 + th^{PCA} \tag{1.16}$$

where x_{proj} defines the projection of the test pattern x onto the subspace as:

$$x_{proj} = W(W^T W)^{-1} W^T x \tag{1.17}$$

and th^{PCA} is the threshold value predefined during the training for accepting or rejecting the data.

This approach has been applied by (Shyu et al., 2003) for analyzing Transmission Control Protocol (TCP) packets for anomaly intrusion detection. They proposed to reduce the dimensionality of the data from 34 original features to 5 major components.

1.2.3.3. Auto-Associative Neural Networks

Neural networks are composed of interconnected processing units arranged in one or several layers that can be used to implement a complex functional mapping between input and output variables. The weights of the neural network are adjusted using training samples so that an error function would be minimized over the training set. The basic design of the Auto-Associative Neural Networks (AANN) is termed ‘‘bottleneck’’. This design assumes that a sample represented in an m-dimensional space is mapped to fewer dimensions and then reproduced for testing the reproduction ability of the model. Usually, an AANN is composed of three layers having in inputs, out outputs and hi neurons on the hidden layer, where hi < in. The AANN is then trained using the standard back-propagation algorithm to learn the identity function over the training set. This design has been used successively by Cottrell and (Zipser et al., 1988) to produce a compression algorithm and (Japkowicz et al., 1995) for novelty detection.

Let n training samples of the target class defined as a set $S = \{x_1, \dots, x_n\}$, the AANN is trained on each sample in order to produce an identity function f that assigns for each input $x_i \in \mathbb{R}^p$ an output $f(x_i) \in \mathbb{R}^p$, $i = 1, \dots, N$ taking ideally the following form:

$$f(x_i) = x_i \tag{1.18}$$

The principle of the AANN is to adjust its weights according to the reconstruction error, which is defined as the absolute distance between output and its corresponding input. Formally, the reconstruction error is defined as:

$$Er(x) = |f(x) - x| \tag{1.19}$$

Such that $x \in S$

A test sample x may either be rejected or accepted according to the threshold value th^{AANN} defined in the training step, as follows:

$$f^{AANN}(x) = -Er(x) + th^{AANN} \tag{1.20}$$

Mainly, the OCC are dedicated to solve one-class classification problems. However, their extension to the multi-class classification is based on training each class on its respective OCC for a defined set of classes $C = \{c_1, \dots, c_m\}$, where m defines the number of classes. A test sample is assigned to the corresponding class when the maximum response is performed.

1.3. Classifier combination approaches

Classifiers combinations become more active and challenging research topic, due to its crucial role for getting more accurate classification accuracy. Thus, the classifiers combination benefits and architectures are presented in the current section.

1.3.1. Benefits of classifier combination

“No Free Lunch” theorems have shown that there is a single classifier that can be considered optimal for all problems (Wolpert, 1996). In addition, there is no clear guideline to choose a set of learning methods and also about the details of how the classification algorithm behaves. Therefore, in practical pattern classification tasks it is difficult to find an appropriate single classifier.

The choice of a single classifier trained with a limited (size or quality) dataset can make the design even more difficult. In this case, selecting the best current classifier can led to the choice of the worst classifier for future data. Especially, when the data used to learn was not sufficiently representative in order to classify properly new objects, the test set provides just apparent errors that differ from true errors. This common situation, where small and not representative data is used as an input to a classifier, can led to difficulties when one must select from a set of possible methods.

According to Dietterich, (2000), there are three main motivations are possible for combining classifiers; statistical, representational and the computational motivation:

- **Statistical (or worst case) motivation**: it is possible to avoid the worst classifier by averaging several classifiers. It was confirmed theoretically by Fumera and Roli, (2005). This simple combination was demonstrated to be efficient in many applications. However, there is no guarantee, that the combination will perform better than the best classifier.
- **Representational (or best case) motivation:** under particular situations, fusion of multiple classifiers can improve the performance of the best individual classifier. It happens when the optimal classifier for a problem is outside the considered “classifier space”. There are many experimental evidences that it is possible if the classifiers in an ensemble makes different errors. This assumption has a theoretical support in some cases when linear combination is performed.
- **Computational motivation:** some algorithms perform an optimization task in order to learn and suffer from local minima. Algorithms such as the back propagation for neural networks are initialized randomly in order to avoid locally optimum solutions. In this case it is a difficult task to find the best classifier, and it is often used several (hundreds or even thousands) initializations in order to find a presumable optimal classifier. Combination of such classifiers showed to stabilize and improve the best single classifier result (Breve et al., 2007).

Other motivations are related to applications, which can naturally use a set of learners such as in sensor fusion.

1.3.2. Classifier combination architecture

There are three main classifier combination architectures, which are parallel, serial and hybrid.

1.3.2.1. Parallel

According to representation showed in figure 1.2, the set of classifiers are trained in parallel, and their output are combined afterwards to give the final decision. More general methodologies were developed using this architecture because it is simpler and easy to analyze.

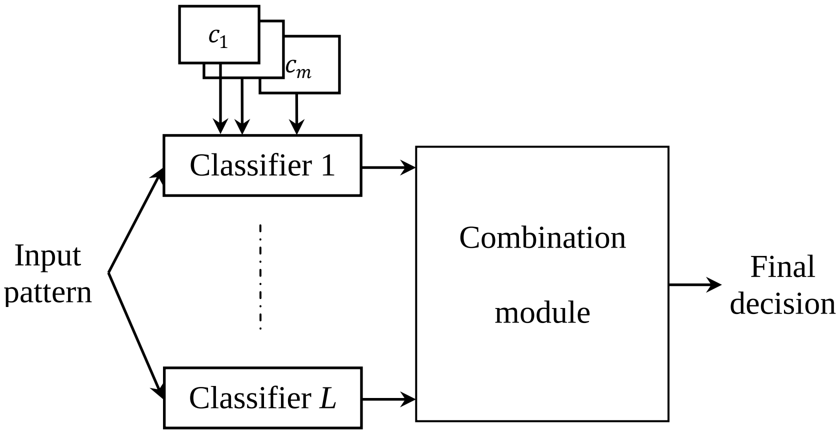


Figure 1. 2: Parallel classifiers combination

1.3.2.2. Serial

As it is depicted in figure 1.3, in the serial approach a primary classifier is used, and when it is not able to classify some new pattern by rejecting it or when confirming a correct classification.

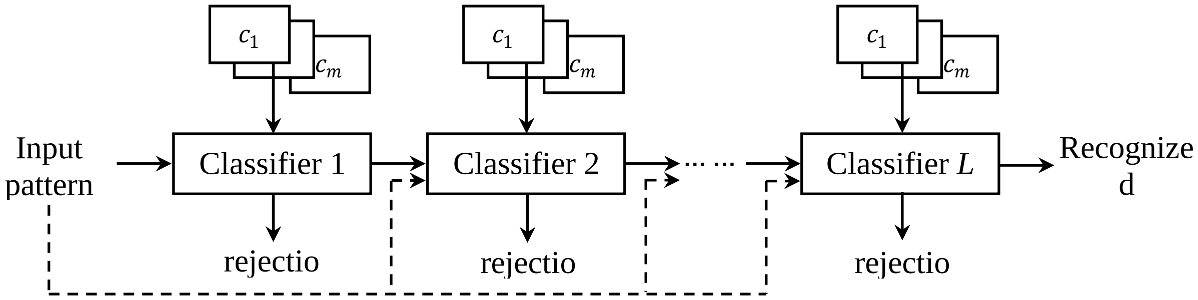


Figure 1. 3: Serail classifiers combination.

A second classifier is used that is trained in order to be accurate on the errors of the previous classifier. A third and fourth classifiers can be used and so on. Sequential methods are often application specific, and are also useful in the context of on-line learning.

1.3.2.3. Hybrid

A hybrid approach combines both sequential and parallel architectures in order to take full advantage of each used classifier. figure 1.4 shows an example of hybrid combination in which a classifier in series is combined with two classifiers in parallel. This kind of approach allows generating many cooperation schemes that can become promptly complex to optimize. It illustrates the two aspects of the combination: on the one hand reducing the set of possible classes, on the other hand searching a consensus between classifiers in order to achieve a single decision.

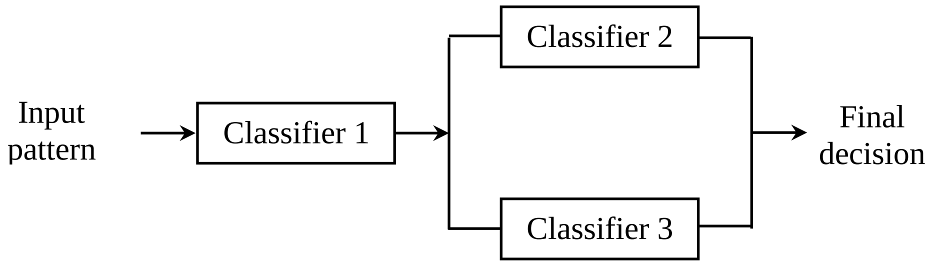


Figure 1. 4: Hybrid classifiers combination.

1.3.3. Classifier combination level

There are many methods for classifiers combination, which are generally classified into three categories according to the level of information provided by the classifier (Kencheva, 2004).

1.3.3.1. Class level

Each classifier produces a class label. At the abstract level, there is no information about the certainty of the guessed labels, nor are any alternative labels suggested. By definition, any classifier is capable of producing a label for test object, so the abstract level is the most universal one.

1.3.3.2. Rank level

The output of each classifier is a subset of labels, with the alternatives ranked in order of plausibility of being the correct label. Rank level based classifier is especially suitable for problems with a large number of classes, for example, character, face, speaker recognition, and so on.

1.3.3.3. Measure level

In the measure level, each classifier produces a m-dimensional vector. The output of the classifier is a vector of measures (normalized or not), which may be a distance, a posterior probability, a confidence value, a match score, belief function, a possibility, credibility or a fuzzy measure, etc.

1.4. Review and taxonomy of OCC ensembles

During the last years, ensemble methods are becoming an active research topic for combining the results produced by a number of weak learning systems to achieve better recognition performance. Various approaches are available for designing an ensemble classifier from individual ones. A survey on the design of multiple classifier systems can be found in (Wozniak et al., 2014). Furthermore, it has been demonstrated that the existing classifier combination strategies can also be used in OCCs. Indeed, OCCs ensemble has been explored to deal with variety of applications. In the following, we propose a taxonomy of OCCs ensemble for which we distinguish single OCC ensemble designed for solving one class problem or novelty detection, multiple OCCs designed for multi-class classification and multiple single OCC ensemble designed for multi-class classification, which defines the hybrid OCC approach.

1.4.1. Single OCC ensemble for novelty detection

OCC aims to distinguish normal object from outliers ones. Similar to standard classifiers, single OCC hardly overfits the data distribution perfectly. Therefore, combining different OCCs via achieving a single OCC ensemble incorporating L different OCCs as it is depicted in figure 1.5 may provide better anomaly detection performance. Indeed, the benefits of single OCC ensemble have been judged in various fields.

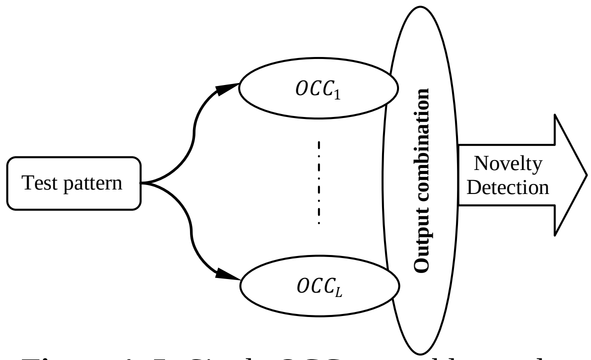


Figure 1. 5: Single OCC ensemble topology.

Single OCC ensemble has been firstly introduced by Tax and Duin, (2001) in which they proved that combining OCCs trained by different feature spaces is very useful. Five methods have been used for the output combination, which are mean vote, mean weighted vote, product of the weighted votes, mean of the estimated probabilities and product combination of the estimated probabilities. The performance of the five methods is compared for a handwritten digit recognition problem using Receiver-Operating Characteristic (ROC) curves in addition to the area under the ROC curve metric. In most situations, the product combination rule gives the best results. In another work, Lai et al., (2002) investigate the single OCC ensemble for image database retrieval through combining Support Vector Data Description (SVDD) classifiers which improve significantly the retrieval precision. Through the well success of single OCC ensemble, Spinosa and de Carvalho, (2004) apply combined OCCs to detect novelty in gene expression data. Experiments were performed with 2 original datasets and a set of 5 classifiers: Parzen Window, K-Nearest Neighbor (K-NN), K-Means, Self Organization Map (SOM) and PCA. Results show that the classification robustness is increased with the use of majority voting combining approach. Nanni, (2006) followed this successful path via investigating the single OCC ensemble for online signature verification. He studied combining several OCCs using the random subspace method. Therefore, new training sets are generated by selecting features randomly which are used to train the selected OCCs: Gaussian model description, Mixture of Gaussian Descriptions, Nearest Neighbor, PCA Description (PCAD), Linear Programming Description (LPD), SVDD and Parzen Window Classifier. Finely, these classifiers are combined through the Max rule. It has been shown that combining different classification methods allows reducing the error rate. Indeed, the best combination method is achieved when using LPD and PCAD.

The benefit of single OCC ensemble has also been considered for information retrieval as reported by Tu et al., (2006), which proposed to extend the one-class bottleneck method for information retrieval by introducing bagging ensemble learning. The proposed ensemble emphasizes different parts of the data and results from different parameter settings are aggregated through Borda method to give a final ranking. Experimental results show improvements in image retrieval applications.

Perdisci and Lee, (2006) proposed single OC-SVM ensemble to construct a payload-based anomaly Intrusion Detection Systems (IDS). Their proposed technique allowed obtaining descriptions of the payload in different feature spaces by constructing several OC-SVMs, which are combined through a simple majority voting rule. Experimental results show that the combination of the obtained classifiers improves both the detection accuracy and the hardness of evasion with respect to other recently proposed payload-based anomaly IDS.

All previous combination methods use fixed rules for combining OCCs. Hence, Gesu and Bosco, (2007) proposed a new method for combining one-class Fuzzy K-Nearest Neighbor (FKNN) classifiers attempting to achieve a robust single OCC ensemble. Their proposed method is based on a genetic algorithm which leads to calculate the appropriate weight assigned to each used FKNN classifier. Results are carried out on two categorical datasets and show that whenever the optimal parameters are found, fuzzy combination based on weighted average of OCCs may improve the overall recognition rate. The single OCC ensemble has been also investigated at the field of biometric combination, as reported by (Bergani et al., (2008). They proposed a system that can be employed at the level of feature selection, score-matching or decision making. Their proposed system has been experimentally evaluated on NIST biometric set for which the weighted sum rule has achieved the best results with the min-max normalization against other normalization methods.

Due to the wide applicability of OCCs for the instruction detection problem, Cabrera et al., (2008) study the problem of distributed intrusion detection in Mobile Ad-Hoc Networks via the application of single OCC ensemble method with special characteristics. Averaging and median schemes are used to combine the anomaly detectors where the averaging scheme performs better than the median one. In a related work, Giacinto et al., (2008) proposed a single OCC ensemble for intrusion detection in computer networks. Different OCCs are used to validate their proposed approach. Fixed rules are the used combiners for aggregating OCCs trained in different feature spaces. The ν-Support Vector Clustering (SVC) shows its superiority against other classifiers by providing the best results in most cases.

Other works have explored the single OCC ensemble to reduce the influence of noisy data for the OCCs. Indeed, Shieh and Kamm, (2009) proposed a single OCC ensemble for combining OC-SVM using bagging method to reduce the sensitivity of the classifier. In fact, even with optimal parameter selection, the OC-SVM can be sensitive to overfitting in the presence of noise. Hence, bagging method is proposed as an alternative solution for reducing the noise influence and prevent overfitting. The achieved classifiers are combined through the majority voting rule to make the final decision. An improved performance of the bagging OC-SVM on both simulated and real world data was achieved. In similar purpose, Segui et al., (2010) proposed two bagging strategies for minimum spanning tree class descriptor attempting to reduce the influence of outliers in training data. They showed that the combined classifiers through fixed rule improve the performance on both real and artificially contaminated data. Always by means of single OCC ensemble for enhancing the performance of novelty detection, Cheplygina and Tax, (2011) proposed a new concept through investigating the effect of pruning on random subspace ensembles on the Gaussian, Nearest Neighbor, and k-Means OCCs. Fixed rules have been used to achieve their proposed ensemble. Experiments show that pruning improves the predictability of the outcomes, but does not always improve the classification performance. However, pruning using the area under the ROC curve shows a significant improvement over the standard ensemble. Authors suggest that combining a few accurate classifiers may be more beneficial than combining all available ones. On the other hand, pruning can only be successful if an appropriate evaluation criterion is selected. (Cyganek, 2012) proposed OC-SVM ensembles for image segmentation and classification. The ensemble is created via clustering the input data and train OC-SVMs by generated clusters. In the operation step, the max rule is used for assigning a point as belonging to the class for which a response of a corresponding weighted OC-SVM. Recently, Krawczyk and Wozniak, (2014 a) studied the effect of the used distance measures for single OCC ensembles, through examining seven different distance measures for five different classifier fusion blocks. Results show that switching to a different distance metric may lead to significant improvement of overall accuracy for the single OCC ensemble.

1.4.2. Multiple OCCs for multi-class classification

The extension of OCCs for multi-class classification has attracted much attention for many researchers for solving the pattern recognition problems. Indeed, the multi-class implementation based on OCCs offers advantages against the usual implementations based on bi-class or multi-class classifiers (Duda and Hart, 2001). Indeed, adding new classes to the multi-class classification does not require retraining the used OCCs for a second time.

As it is depicted in figure 1.6, the extension of OCC for the multi-class classification is achieved by m OCCs such that m defines the number of the predefined classes. During the recognition phase, the outputs of the OCCs are combined to assign a class label for each test pattern.

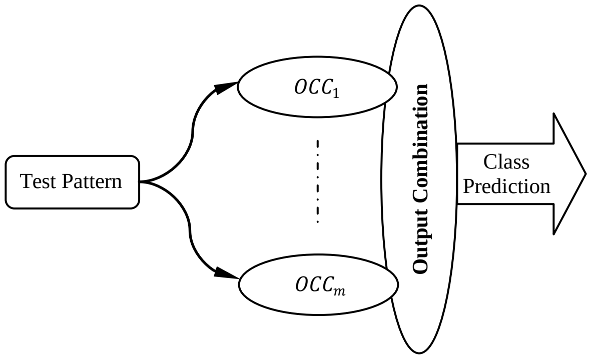


Figure 1. 6: Multiple OCCs for multi-class implementation topology

In this context, Sun and Huang, (2003) addressed the drawbacks of implementing Bi-SVM for multi-class classification as time consuming and performance degradation, especially when dealing with high number of classes. Thus, they proposed the use of SVC classifier which learns the model of the *m* classes separately. However, this method can suffer from a scaling problem which is solved by combining outputs of SVCs via a two-layered neural network as a combination model. Compared against to the Bi-SVM, their approach can reduce the computational complexity and get better performance at the same time on the case of multi-class classification. Besides, there is no requirement of having the same kernel function and parameters for all classes, which can guarantee to get better boundaries for each class. Consequently, better classification results are achieved. However, it is worthwhile to point out that the output combination of the neural network achieves a closed system. Indeed, adding new classes requires designing a new architecture of the neural network classifier. In related work, Munroe and Madden, (2005) extend the idea of one-class K-NN for vehicles recognition by means of a feature set extracted from their frontal view. In order to enhance the performance, (Sachs et al., 2006) proposed two strategies for combining OC-SVMs for multi-class classification of bioacoustic time series. Experimental results show that the nearest support vectors strategy performs better than the nearest center one as well as than k-nearest neighbor and learning vector quantization network classifiers. In similar concept, (Ban and Abe, 2006) address the problem of how to implement a multi-class classifier by multiple OCCs. Firstly, OCCs are trained for each class and then a decision function is formulated based on minimum distance rules. Two types of OCCs have been explored: the SVDD and Kernel Principle Component Analysis (KPCA) based method. Experiments on some benchmark datasets show that the proposed decision function with carefully tuned parameters have comparable generalization ability with Bi-SVM while having some other advantages.

Lee et al., (2007) proposed a novel method by integrating *m* OC-SVMs with various discriminant functions for multi-class classification task. Three discriminant functions have been proposed and experimental results have shown that similarity measure and Z-score measure outperform the other methods including nearest center and nearest support vector (Sachs et al., 2006). The problem of the discriminant functions has also been addressed by (Rabaoui et al., 2007) via proposing the use of discriminative method based on OC-SVM, which proposed a logarithmic function in order to classify a set of sounds into predefined classes. Experiments show that the multi-class implantation based on the OC-SVM and the proposed discriminant function outperform the conventional HMM classifier for sound classification. Boehm et al., (2011) explored the one-class neural network classifier for recognizing cognitive activities from brain activation data. It has been shown that a joint use of the feature selection genetic algorithm and the one-class neural network can lead to increase the accuracy of the classification closely to that obtainable from bi-class methods. Recently, Wilk and Wozniak, (2012) addressed the problem of OCCs combination to achieve multi-class implementation. While different approaches have been used for the combination step such as max rule, nearest center, nearest support vectors, normalization of outputs and neural network classifier, the authors proposed fuzzy combiners to achieve more accurate system. They mentioned that the combination methods of OCCs performed relatively better in comparison to their binary counterparts

1.4.3. Hybrid OCC ensemble

Hybrid OCC ensemble is defined as a multi-class implementation based on single OCC ensembles. It is designed for attempting to enhance the recognition performance and system robustness of the OCC multi-class implementation. As it is depicted in figure 1.7, the test pattern is assigned to one of the predefined classes according to two steps. First, a single OCC ensemble per class is constructed from which outputs are combined through a predefined combination rule. Second, the outputs of multiple single OCC ensembles are combined through another combination rule for predicting the class label.

Owing to its offered advantages than conventional multi-class classifiers, hybrid OCC ensemble finds its applicability in various applications. Indeed, in addition to the significant improvement of the performance, the hybrid OCC ensemble preserves the property of an open system since it is possible to add new classes without retraining all single OCC ensembles. Hence, this scheme has firstly been explored by Juszczak and Duin, (2004) for classifying missing data in multi-class problems. In their proposed system, a single Parzen OCC ensemble is trained for each existing class in the training set. Each ensemble contains one classifier for each feature. During the classification step, only the available features are classified, which are combined using a fixed rule. In comparison against the standard methods, their proposed method offers two advantages. Firstly, it requires fewer classifiers to be trained. Secondly, in the classification step, it does not require retraining the system whenever missing feature values occur. Later, Goh et al., (2005) pointed that the bi-class classifier is susceptible to the problems of noisy and imbalanced training data (Wu and Chang, 2003). Hence, they proposed the use of OC-SVM to estimate the support of individual classes leading to avoid these problems. Moreover, bagging scheme has been proposed to reduce class prediction variance. Finely, the overall class prediction is the result of majority voting of the several multi-class implementations bags. Furthermore, they explored the mentioned advantages of OCC for constructing dynamic ensemble which can be used for new class discovery. Experimental results showed the effectiveness of their proposed system. In a related work, (Muñoz-Marí et al., 2007) demonstrate that using a simple combination rule (e.g. average or product) on OCCs trained on different feature sets are able to improve the classification accuracy for classifying image remote sensing. They proved that the ensemble composed of the support vector data description achieves better results in comparison with mixture of Gaussian ensemble.

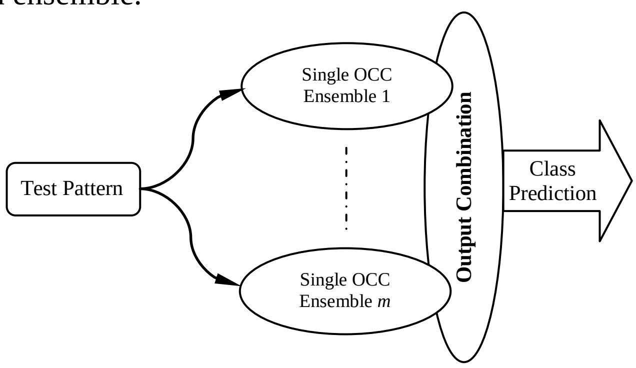


Figure 1. 7: Hybrid OCC ensemble topology

AdaBoost.MK algorithm has been also investigated for hybrid OCC ensemble by Yeh et al. (2009) [9], which proposed a combination scheme by using conjointly AdaBoost.MK and One Class Support Vector Machine (OC-SVM) with a well-designed discriminant function for solving the multi-class classification problems. Finally, the classifiers returned by AdaBoost.MK. are aggregated through the sum combination rule. In another work, Krawczyk and Wozniak (2012b) follow the path of using the diversity measure that has been widely explored for conventional multi-class classifiers to achieve complimentary systems. In their proposed system, OC-SVM ensembles trained on different training sets are performed. Then, a diversity measure is applied on the constructed ensembles in order to select the suitable ones for each class. Finally, the selected classifiers are combined using (Error Correcting Output Code) ECOC (Dietterich and Bakiri, 1995) and (Decision Template) DT (Kuncheva et al., 2001) strategies to generate a robust system.

More recently, a great interest can be observed for combining multiple single OCC ensembles. Indeed, Abbas et al., (2013) used the Dezert-Smarandache theory to achieve OC-SVM ensemble. The proposed system relies on training the OC-SVM on different feature sets of handwritten digits to create diverse classifiers. The DSmT shows its superiority in term of performance versus the sum rule. However, their proposed scheme is considered as a closed system and therefore violates the best advantage of using OCCs as an open multi-class system. Indeed, adding new classes require updating all parameters and retraining the combination model. Krawczyk and Filipczuk, (2014) proposed an efficient medical decision support framework that allows distinguishing between benign, malignant and fibro adenoma cases, based on combining OCC trained on different features and the ECOC combination strategy. Experimental evaluation shows superiority of their proposed system against some state of the art systems. In similar concept, hybrid OCC ensemble has been also investigated for medical image classification (Zhang et al., 2014). The ensemble consists of one-class KPCA models trained on different features generated from each image class. Besides, a new product combination rule was proposed. The effectiveness of their proposed classification scheme was verified using a breast cancer biopsy image dataset and a 3D optical coherence tomography retinal image set. The proposed classification scheme obtained competitive results on the two medical image sets in comparison against the stat of the art systems. Recently, in a related work (Krawczyk et al., 2014) presented a new method to design the hybrid OCC ensemble based on a single OCC ensemble. The main idea is to create single OCC ensemble for each class based on feature space partitioning. The combined classifiers are trained on the basis of clusters that lead to make use of individual classifier strengths. Experiments carried out using a wide range of benchmark datasets, prove the validity and the flexibility of the proposed framework to work with different clustering algorithms. In order to improve their system, Cyganek and Krawczyk, (2015) proposed to split data using the nonnegative matrix factorization algorithm with sparse constraints. This framework relies on splitting the input data into compact and consistent clusters as well as automatic determination of the cluster's number. The proposed method shows high accuracy and fast classification.

1.5. Classifier combination methods

The combination method is the successful key of a multiple classifier system. In literature, various methods have been developed leading to maximize the benefits from using multiple systems. However, the combination of hybrid OCC ensembles sill await for proper attention, since, simple fixed rules are mainly explored. Thus, in this thesis we attempt to explore more adapted weighting combinations and the Fuzzy Integral (FI) operators for the hybrid OCC ensembles. In the following we present brief review of fixed rules as well as most known FI operators.

1.5.1. Fixed rules

Fixed rules are considered as linear combiners that have no extra parameters to be tuned. Thus, the ensemble members are considered similar in terms of performance and competence. The most used rules are summarized in table 1.1 for combining the a posteriori probabilities $P_i\,(c_j/x)$ yield by L classifiers (Kuncheva, 2004).

Table 1.1: Mathematical equations of fixed rules for class prediction

Combination rule	Class label
Average	$y(x) = \underset{j=1}{\overset{m}{argmax}} \left(1/L \sum_{i=1}^{L} P_i\,(c_j/x) \right)$
Product	$y(x) = \underset{j=1}{\overset{m}{argmax}} \left(\prod_{i=1}^{L} P_i\,(c_j/x) \right)$
Maximum	$y(x) = \underset{j=1}{\overset{m}{argmax}} \left(\underset{i=1}{\overset{L}{max}} \left(P_i\,(c_j/x)\right) \right)$
Minimum	$y(x) = \underset{j=1}{\overset{m}{argmax}} \left(\underset{i=1}{\overset{L}{min}} \left(P_i\,(c_j/x)\right) \right)$

1.5.2. Fuzzy integral

Fuzzy integrals are non-linear combiners defined with respect to fuzzy measures. Therefore, the main advantage of FI is its ability to combine the objective evidences in the form of expert decisions taking into account subjective evaluation of their competence expressed by the fuzzy measure. In this section, we review the main properties of fuzzy measures in addition to the used fuzzy integral operators.

1.5.2.1. Fuzzy measure

Let $Z = \{z_1, \ldots, z_L\}$ defining a finite set of elements, a set function: $2^Z \rightarrow [0, 1]$ is called fuzzy measure if it verifies the following properties (Cho and Kim, 1995; Cho 1995):

1. $g(\phi) = 0,\ g(Z) = 1$
2. If $A, B \subset 2^Z$ and $A \subset B$, then, $g(A) \leq g(B)$
3. If $A_n \subset 2^Z$ for $1 \leq n \leq \infty$ and the sequence $\{A_n\}$ is monotone in the sense of inclusion then,

$$\lim_{n\rightarrow\infty} g(A_n) = g\left(\lim_{n\rightarrow\infty} A_n\right) \tag{1.21}$$

It is obvious that the fuzzy measure is not necessary additive, therefore, if $A, B \subset Z$, and $A \cap B = \phi$, then:

$$g(A \cup B) \neq g(A) + g(B) \tag{1.22}$$

According to this inequality, Sugeno (1977) introduced the decomposable so called λ-fuzzy measure satisfying the following additional property for $\lambda > -1$:

$$g(A \cup B) = g(A) + g(B) + \lambda g(A)g(B) \tag{1.23}$$

Let consider $Z = \{z_1, \ldots, z_L\}$ as the set of information sources to be aggregating for attempting to achieve better performance, each information source z_i is then associated to a density measure $g^i = g(\{z_i\})$, which represents generally the performance of the single source alone. The set of objective evidence, denoted $h(z_i)$, is rearranged from maximum to the minimum value as $h(z_1) > \cdots > h(z_L)$ as well as their corresponding density measures g^i. Then according to this order, the λ-fuzzy measure $g(A_i)$ of the new sequence $A_i = \{z_1, \ldots, z_i\}$ can be computed recursively as follows:

$$g(A_1) = g(\{z_1\}) = g^1 \tag{1.24}$$

$$g(A_i) = g(A_{i-1}) + g^i + \lambda g(A_{i-1})g^i,\ 2 \leq i \leq L \tag{1.25}$$

The value of λ is deduced by solving the equation $g(Z) = 1$, which corresponds to resolve the following equation:

$$\lambda + 1 = \prod_{i=1}^{L}\left(1 + \lambda g^i\right) \tag{1.26}$$

This value is obtained as the unique real root greater than -1 and not equal to zero.

1.5.2.2. *Fuzzy integral operators*

Sugeno integral I_s is the first FI operator defined for aggregating a function $h : Z \to [0, 1]$ with respect to a fuzzy measure g over Z. It is computed as:

$$I_S = max_{i=1}^{L}\left[min\big(h(z_i), g(A_i)\big)\right] \tag{1.27}$$

This integral represents the simplest operator for FI combination. Hence, it has been extended through the definition of the Choquet integral. Its discrete formulation for a function $h : Z \to R^+$ with respect to fuzzy measure g is defined as:

$$I_C = \sum_{i=1}^{L}\{h(z_i) - h(z_{i-1})\}\, g(A_i) \tag{1.28}$$

Where indices i have been permuted so that $0 \le h(z_1) \le \cdots \le h(z_L) \le 1$ and $h(z_0) = 0$.

Attempting to enhance more efficiently the FI, other efficient operators have been proposed such as Ordered Weighted Averaging **(**OWA). In the following, we recall two commonly used OWA namely OWA-AND and OWA-OR.

The OWA-AND requires the calculation new of evidences $\hat{h}(z_l)$ defined as:

$$\hat{h}(z_l) = \frac{(1-\alpha)}{l}\sum_{i=1}^{l} h(z_i) + \alpha\, min_{z_l \in Z}\,\{h(z_l)\} \tag{1.29}$$

These new evidences are then used into Sugeno integral, which is termed I_{S-AND}. In addition, we propose in this work to use these new evidences into the Choquet integral termed I_{C-AND} in order to evaluate its performance

In contrast, the OWA-OR is performed using the same new evidences through the new decision function defined as follows:

$$I_{OR} = \frac{1-\beta}{2^L}\sum_{i=1}^{L} min\left(\hat{h}(z_l), g(A_i)\right) + \beta\, max_{i=1}^{L}\left[min\left(\hat{h}(z_l), g(A_i)\right)\right] \tag{1.30}$$

Both operators need tuning parameters α and β in the unit interval in order to achieve better results than the Sugeno and Choquet operators (Cho, 1995):. In the following, the five different operators for FI are termed Sugeno (I_S), Choquet (I_C), S-AND (I_{S-AND}), C- AND (I_{C-AND}) and OR (I_{OR})

1.6. Conclusion

In this chapter, we have presented an overview of the most known OCCs in addition to the classifier combination approaches. We have first recalled the mathematical foundation of the most known OCCs in the literature. Furthermore, the state-of-the-art that is related with OCC ensembles was studied with more details and we proposed a taxonomy of OCC ensembles.

Moreover, classifier combination approach is presented, which can be performed through three combination schemes, namely, sequential combination scheme, parallel combination scheme and the hybrid combination scheme. In this thesis, the parallel classifier combination is adopted and combined in the measure level. This later is performed through standard fixed rules and more efficient combination based on FI, which will be explored in some applications in the present thesis for OCC ensembles improvement.

The next chapters are dedicated to present the use of the OCC ensembles improvement for pattern classification applications in general as well as for specific handwriting recognition applications.

Chapter 2

Diverse One-Class Classifiers Combination for Multi-Class Pattern Classification

Abstract

One-Class Classifier has been widely used for its ability to learn without counterexamples. Its extension for multi-class implementation offers an open scheme which allows easily adding new classes. However, using OCCs for the multi-class implementation usually achieves less accuracy than the usual multi-class implementations. Hence, in order to improve the accuracy and keep the offered advantage, the suitable approach consists to combine different classifiers. Thus, in this chapter we proposes a study of combining different types of OCC for multi-class classification by means a new Dynamic Weighted Average (DWA) combination rule. Experimental results conducted on several real-world datasets prove the effective use of the proposed approach where the DWA rule achieves the best results against fixed rules as well as the decision template. Furthermore, comparison of the proposed open classification system against a standard open classifier based on K-NN shows the superiority of the proposed system

2.1. Introduction

The OCC has been successfully employed in many applications as image retrieval (Kwang-Kyu, 2007), automated document retrieval and classification (Manevitz and Yousef, 2007), combining different biometric traits (Bergani et al., 2009), anomaly detection in wireless sensor networks (Moshtaghi et al., 2011), novelty detection in wildlife scenes (Yong et al., 2012) and image segmentation and classification (Cyganek, 2012).

In the last decades, OCC has attracted much attention for many researchers leading to use it for solving the multi-class classification problem (Goh et al., 2005; Ban et al., 2006; Yeh et al., 2009; Rabaoui et al., 2007; Boehm et al., 2011 ; Seokho et al., 2015). Those are motivated by the advantages offered by the OCC against the usual implementations based on bi-class or multi-class classifiers (Duda et al., 2001). Indeed, adding new classes does not require retraining the used OCC for a second time. In addition, the OCC offers less computational cost in terms of training time and memory space against some multi-class implementations (Yeh et al., 2009) such as One-Against-One (OAO) and One-Against-All (OAA) (Hsu and Lin, 2003).

However, using a single type of OCC for the multi-class implementation usually achieves less accuracy than the usual multi-class implementations (Boehm et al., 2011). Furthermore, due to the high diversity of existing OCCs (Tax and Duin, 2004), choosing a specific classifier for various applications is a difficult task (Kuncheva, 2004). Therefore, a multiple classifier system (MCS) is more suitable since it can produce a better system in terms of robustness and accuracy (Duin, 2002). In addition, it allows keeping the offered advantage for achieving an extensible multi-class system. In this case, the most difficult problem is finding the appropriate classifiers and adapted combination rule for the addressed problem.

In order to perform the combination, an ensemble of diverse classifier must be created for which different ways are possible. The most popular ways are based on different initialization, different parameter choices, different architectures, different classifiers, different training sets or different feature sets (Duin, 2002).

However, when combining OCCs for multi-class implementation as it has been reported in the first chapter, most hybrid OCC ensembles explore different training sets or different feature sets to yield diverse systems. Indeed, (Juszczak and Duin, 2004) used hybrid Parzen OCC ensembles trained on different feature sets. In a related works, (Muñoz-Marí et al., 2007; Abbas et al., 2013; Krawczyk and Filipczuk, 2014) explored same approach for different applications. More recently, (Krawczyk and Wozniak, 2014) use the OC-SVM ensemble trained on different training sets.

Nevertheless, different feature sets are not always available or complementary. Moreover, in some applications such as handwritten signature verification or identification, training samples are often reduced, which does not allow generating different training sets.

Hence, in this chapter a study of combining different types of OCC for multi-class implementation is presented *via* proposing more suitable combination rule. Usually, OCCs are combined using simple combination strategies such as fixed rules (Juszczak, R.P.W. Duin, 2002; Muñoz-Marí et al., 2007), ECOC and DT strategies (Krawczyk and Filipczuk, 2014; Krawczyk and Wozniak, 2014). In this chapter, an alternative approach is proposed based on a Dynamic Weighted Average rule (Hadjadji et al., 2014a; Hadjadji et al., 2017a) instead of the popular Static Weighted Average (SWA) (Verikas et al., 1999; Al-Ani and Deriche, 2002; Wang et al., 2002), leading to measure the importance of the used classifiers through calculating their suitable weights for each test sample.

The remaining of this chapter is organized as follows. Section 2 describes the proposed MCS using an ensemble of diverse OCCs and the combination rule based on various combination rules. Section 3 presents experimental results conducted on different datasets in order to prove the effective use of the proposed MCS. Finally, the conclusion and future work are provided in the last section.

2.2. Combination of one-class classifiers for multi-class implementation

The basic structure of the proposed MCS based on OCCs is depicted in figure 2.1. The achieved MCS is a collection of multiple OCCs, each one defines a specific class which is composed of *L OCCs*, operating in parallel at the same data. Their normalized outputs are aggregated through a combination rule. A description of each stage of the MCS is given in the following sections.

2.2.1. Normalization of OCC outputs

Several combination rules are possible to achieve the MCS, but all these rules need a unique interpretation of the outputs generated by the different classifiers for each test sample x. Hence, the output normalization for each classifier is required for performing correctly the combination. In this way, a straightforward approach consists to transform the classifier outputs into posterior probabilities. Thus, we propose to use the softmax normalization method (kuncheva, 2004) for its simplicity and effectiveness to map the i^{th} classifier output $f_j^{OCC\ i}$ of j^{th} class defined as c_j to posterior probabilities $P_i(c_j/x)$ ranging between [0,1] as follows:

$$P_i(c_j/x) = \frac{exp\left(f_j^{OCC\ i}(x)\right)+1}{\sum_{j=1}^{m} exp\left(f_j^{OCC\ i}(x)\right)+1} \tag{2.1}$$

2.2.2. Combination rules

The combination rule is the key for achieving an enhanced MCS. Thus, we use fixed

(Juszczak, R.P.W. Duin, 2002; Muñoz-Marí et al., 2007), and trained rules including DT (Krawczyk and Filipczuk, 2014; Krawczyk and Wozniak, 2014). and Static Weighted Average (SWA) (Verikas et al., 1999; Al-Ani and Deriche, 2002; Wang et al., 2002), as a base of comparison to validate the proposed Dynamic Weighted Average (DWA) combination. In the following, the basic as well as the proposed combination rules are described for designing a Multiple Classifier System (MCS).

The fixed combination rules assume that all classifiers have the same importance or competence. However, when using different types of OCC, each classifier has its own significance that should be considered. Hence, trained rules are considered as an alternative approach that allows taking into account the significance of the different used classifiers.

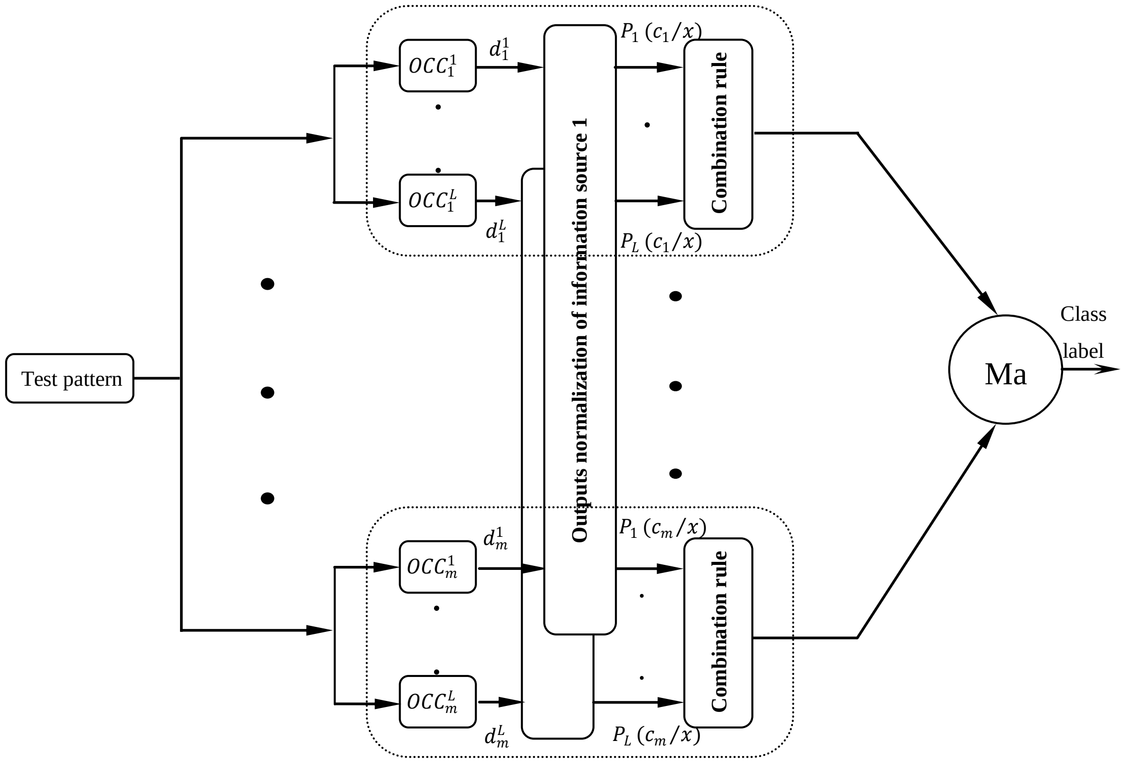


Figure 2.1: The proposed MCS architecture.

DT and ECOC are well known combination strategies for OCC. Indeed, they have been investigated for combining OCC leading to achieve multi-class implementation (Wilk and Wozniak, 2012) and also for combining different multi-class implementations (Krawczyk and Filipczuk, 2014; Krawczyk and Wozniak, 2014). The achieved results are however roughly similar for both combinations. The ECOC is not considered in this work since it relies on aggregating output labels that requires attributing a threshold value for each used classifier. Hence, only combination of continuous-valued outputs strategy is explored and the threshold values are set to zero.

2.2.2.1. Decision template

The idea behind the DT combiner is to remember the most typical decision profile for each class, called the decision template DT_j, which is often deduced by averaging classifier outputs generated from the training samples.

During the testing step, the current Decision Profile $DP(x)$ is compared with the decision template of the different classes DT that are already deduced form the training samples. The matching process can be performed using different distance measures from a set of different ones. Nevertheless, the Euclidean distance has proved its effective use against other measures for performing DT-based classifier combination (Kuncheva, 2003). Furthermore, (Wilk and Wozniak, 2014 a) reported that Euclidean distance is widely used for DT distance criterion. Indeed, it has been successfully investigated for bi-class and one-class classifiers combination for multi-class implementation. Thus, only the Euclidian distance d_E is used in order to assign the evaluated sample x to the closest match as follows:

$$y(x) = \underset{j=1}{\overset{m}{argmax}} \left(1 - d_E\left(DP(x), DT_j\right)\right) \tag{2.2}$$

2.2.2.2. Static weighted average

Generally, the SWA is considered as a successful straightforward combination rule, which has been widely used for combining classifiers (Verikas et al., 1999; Al-Ani and Deriche, 2002; Wang et al., 2002), as a weighted sum and weighted Borda count rules. In addition, it has been also used for multimodal biometric authentication, where weights are assigned to each matcher for biometric fusion. In our case, weights are assigned to each individual classifier according to its importance for achieving the MCS. However, this importance is assumed the same for all test samples. Indeed, the individual classifiers can differ from each other in terms of performance. This difference can be measured using a validation dataset which can be used for selecting the weights to be assigned to the ensemble members.

Thus, this method relies on the Average Error Rates (AERs) of classifiers which are calculated during the validation step. Denote the AER assigned to each classifier i as r_i, $i = 1, \dots, L$ where L defines the number of classifiers used in a ensemble. Then, the weight w_i associated to the classifier i is calculated as follows:

$$w_i = \frac{1/\sum_{i=1}^{L}\frac{1}{r_i}}{r_i} \tag{2.3}$$

Such that $0 \leq w_i \leq 1$ and $\sum_{i=1}^{L} w_i = 1$. The class label $y(x)$ of a test sample x is defined using the following equation:

$$y(x) = \underset{j=1}{\overset{m}{argmax}}\left(\sum_{i=1}^{L} w_i P_i\,(c_j/x)\right) \tag{2.4}$$

2.2.2.3. Dynamic weighted average

The main drawback of the SWA is that all samples are weighted by the same values. This approach is not efficient since each test sample has its suitable classifier and, therefore, more contribution should be given in respect to the more appropriate classifier. Indeed, each classifier allows recognizing well a set of samples which is not well recognized by the others. Therefore, the contribution of each classifier for each test sample must be calculated. In order to overcome this drawback, a DWA is proposed here which calculates the weight assigned to each classifier for each test sample according to the classifier maximum response kept from the training step. More precisely, when the response of the classifier (i.e. the maximum value of all classes) is near to its maximum response relatively to other classifiers, a high contribution is assigned to this one versus the other classifiers. In contrast, when the response of the classifier is less than its maximum response relatively to the other classifiers, the contribution will be minimized. Let $w_i(x)$ be the weight assigned to the i^{th} classifier for a test sample x and denote $op_i^j(x)$ the j^{th} class of i^{th} classifier response and op_i^{max} its maximum response calculated during the training step, the weight assigned to the i^{th} classifier for the test sample x is defined as:

$$w_i(x) = \frac{exp\left(\overset{m}{\underset{j=1}{max}}\left(op_i^j(x)\right)\right)/\,exp\left(op_i^{max}\right)}{\sum_{i=1}^{L} exp\left(\overset{m}{\underset{j=1}{max}}\left(op_i^j(x)\right)\right)/\,exp\left(op_i^{max}\right)} \tag{2.5}$$

Such that, $0 \leq w_i(x) \leq 1$ and $\sum_{i=1}^{L} w_i(x) = 1$. Hence, op_i^{max} is used as a reference of classifier output comparison. In other word, it is used as a control value for estimating the suitable contribution of each one.

Moreover, for each test sample, the proposed DWA allows to favorite the classifier that responds nearly to its maximum response relatively to the other members in the ensemble.

Indeed, the nearest response of the classifier to its maximum response is more certain to produce a correct classification than the others. A test sample x is then assigned to the class label $y(x)$ according to the following equation:

$$y(x) = \underset{j=1}{\overset{m}{argmax}}\left(\sum_{i=1}^{L} w_i(x)\, P_i\,(c_j/x)\right) \tag{2.6}$$

In summary, the proposed DWA is performed according to Algorithm 2.1.

Algorithm 2.1:

Inputs: Test sample x, classifier models OCC_j^i, maximum classifier outputs op_i^{max}

Output: Class label

1: for $i \leftarrow 1$ to L do /* i represents OCC type variable */

2: for $j \leftarrow 1$ to m do /* j represents class variable */

3: Calculate the classifier output $op_i^j\,(x)$

4: end for

5: Compute the normalized output according to Eq. 2.1

6: Compute $w_i(x)$ according to Eq. 2.5

7: end for

8: Combine classifiers for computing the class label via Eq. 2.6

2.3. Experimental results

2.3.1. Dataset description

In order to evaluate the effective use of the proposed MCS, four datasets are selected from ELENA project[1], which represent real applications: Iris, Phoneme, Satimage and Texture. In addition, we use Breast cancer[2], Crab gender[3], Handwritten digit[4], off-line handwritten

[1] ftp.dice.ucl.ac.be in the directory pub/neural-nets/ELENA/databases.

[2] http://mlearn.ics.uci.edu/databases/breast-cancer-wisconsin/.

[3] http://www.stats.ox.ac.uk/pub/PRNN/.

signature datasets from the Center of Excellence for Document Analysis and Recognition (CEDAR) and a set of handwritten Arabic word taken from the well known dataset IFN/ENIT[5] used for recognition purpose. There are a few missing values in the used datasets. In this case, during experiments, these values are replaced by the average of the feature regardless of the class labels as it has been performed in (Kuncheva, 2003). All these datasets are summarized in Table 2.1.

Table 2.1: Datasets used for evaluating the proposed MCS.

Dataset	# Classes	# Features	# samples
Phoneme	2	5	5404
Iris	3	4	150
Texture	11	40	5500
Satimage	6	36	6435
Breast Cancer	2	9	699
Crab Gender	2	6	200
Digits (USPS)	10	86	7287 / 2007
CEDAR	55	48	1320
Arabic word	100	768	12194

2.3.2. Experimental setup

The MCS is composed of five types of OCC which are trained separately on the same feature space using the same training set. However, each classifier has its own parameters which must be carefully tuned. For the OC-NN classifier, no parameters are required to be tuned unlike to other classifiers. For the K-Means and k-Center classifiers, the respective parameters (the number of cluster K and center k) are very sensitive and therefore should be tuned carefully to avoid a misclassification (Dietterich, 2000). Hence, the classifiers are trained with different values and then the optimal values are selected when the best performance of the training dataset is achieved. For PCA an only one parameter must be specified, which is the number of

[4] http://www.gaussianprocess.org/gpml/data/.

[5] www.ifnenit.com.

eigenvectors p. Hence, the classifier is trained with different values of p and then the optimal value is selected when the best reconstruction is achieved from the training dataset. In contrast, the OC-SVM has two parameters (ν,γ), which are fixed for each class during the training step. Hence, different couples of parameters are generated during the training step in order to achieve the best recognition rate. The optimal couple is then selected when both highest recognition rate and the number of support vectors are obtained from the training dataset.

In the experiment step, stratified three-fold cross validation is performed for all datasets except for Digits dataset where the training and testing samples are defined beforehand as mentioned in table 2.1.

2.3.3. Experimental evaluations

2.3.3.1. Performance evaluation

Results for the individual classifiers and MCS with different combination rules conducted on the used datasets are reported in Table 2.2 and 2.3, respectively. Firstly, when comparing the individual classifiers, we can notice that there is no–dominant classifier suitable for all datasets. For instance, the OC-SVM provides the highest results only for Breast Cancer dataset. Secondly, we can also notice that combining individual classifiers allow improving the recognition rates than the best single one for all datasets, except for DT combination where lower results than the best classifier are obtained for dataset having high number of classes (Texture, digits and CEDAR). The DT is used for OCCs multi-class implementation (Sachs et al., 2006) and also for combining the different multi-class implementations (Krawczyk and Filipczuk, 2014; Krawczyk and Wozniak, 2014).. We can clearly observe that the DT achieves poor results for multi-class implementations when the number of classes is large as for CEDAR dataset, which leads to achieve less result against single classifiers that have been implemented without DT.

Besides, when observing carefully the obtained results, we can notice that the DWA rule offers an improved recognition rate whatever the selected dataset. Indeed, it is interesting to note that when using the Crab gender and digits datasets, DWA is the unique combination rule that allows improving the recognition rate. Furthermore, the DWA shows its robustness to deal with large number of classes for CEDAR dataset and provides the best results.

Table 2.2: Classification accuracy (%) of individual one class classifiers

Database	1NN	K-Means	k-Center	SVM	PCA
Phoneme	**78.11±0.48**	70.66±3.72	72.16±2.39	74.65±0.95	74.30±1.36
Iris	96.07±1.96	**96.73±1.13**	94.77±1.13	95.42±4.93	95.41±1.33
Texture	**98.80±0.16**	96.37±0.34	96.73±0.69	98.22±0.22	98.34±0.25
Satimage	84.47±0.79	85.26±0.99	**87.28±0.85**	83.69±1.13	69.77±1.05
Breast Cancer	91.31±0.98	95.01±1.78	95.15±2.35	**95.16±2.76**	88.75±2.85
Crab Gender	**93.10±3.34**	86.70±6.64	88.65±4.39	90.64±4.16	**93.10±3.34**
Digits (USPS)	92.47±0.00	92.27±0.00	91.08±0.00	93.32±0.00	**93.97±0.00**
CEDAR	86.01+0.61	86.75±1.70	85.43+1.95	84.80±0.42	**87.87±0.60**
Arabic word	82.43±0.86	80.36±1.36	80.17±1.86	80.95±0.31	**83.34±0.78**

Table 2.3: Classification accuracy (%) of MCS with different combination rules.

Dataset	Combination rule					
	Average	Max	Prod	DT	SWA	DWA
Phoneme	79.17±0.61	78.69±0.64	79.17±0.61	77.32±0.78	79.07±0.55	**79.52±0.45**
Iris	**98.69±1.13**	96.07±3.39	**98.69±1.13**	97.38±1.13	**98.69±1.13**	**98.69±1.13**
Texture	98.82±0.16	98.29±0.15	98.80±0.10	89.45±1.19	98.83±0.19	**99.49±0.45**
Satimage	86.63±0.60	83.69±1.13	86.63±0.60	87.34±0.14	87.14±0.97	**89.55 ±0.65**
Breast Cancer	96.71±2.03	96.72±1.37	96.71±2.03	96.43±2.36	**96.86±2.20**	**96.86± 2.20**
Crab Gender	92.12±4.44	92.61±3.84	92.12±4.44	92.12±2.12	92.12±4.19	**93.61±4.21**
Digits (USPS)	93.72±0.00	93.47±0.00	93.72±0.00	90.08±0.00	93.72±0.00	**94.02±0.00**
CEDAR	87.76±0.71	84.80±0.42	87.56±0.96	74.94±2.05	87.90±1.21	**89.17±0.59**
Arabic word	84.73±0.43	80.67±0.20	84.75±0.48	68.28±0.93	84.77±0.48	**85.37±0.34**

In order to prove the effective use of the DWA combination rule against the existing combination rules, we use the McNemmar's test (Dietterich, 1998), which allows comparing statistically two systems. More precisely, a contingency table is constructed in order to calculate the p-value as reported in table 2.4.

Table 2.4: 2×2 Contingency table.

n_{00}	n_{01}
n_{10}	n_{11}

- n_{00} is number of samples misclassified by both systems.

- n_{10} is number of samples misclassified by the system I and not by the system II.
- n_{01} is number of samples misclassified by the system II and not by the system I.
- n_{11} is number of samples correctly classified by both systems.

The McNemar's test can provide whether one system is significantly better than another according to the *p*-value. More the *p*-value is smaller, more the difference of accuracy is likely to be more significant. However, when the *p*-value exceeds 0.05, then the hypothesis is considered as null. In this case, both systems perform closely and the difference is too small to decide the superiority of one system than the other. Table 2.5 reports the *p*-value of the DWA rule against the remaining combination rules (Average, Max, Prod, SWA and DT) for all used datasets.

Table 2.5: The p-values of McNemar's test for DWA versus the other combination rules.

Dataset	DWA vs Average	DWA vs Max	DWA vs Product	DWA vs SW	DWA vs DT
Phoneme	$\mathbf{6.66 \pm 2.90 \times 10^{-3}}$	$\mathbf{0.30 \pm 0.97 \times 10^{-3}}$	$\mathbf{6.66 \pm 2.90 \times 10^{-3}}$	$\mathbf{1.30 \pm 0.63 \times 10^{-3}}$	$\mathbf{0.10 \pm 0.80 \times 10^{-3}}$
Iris	NA	$1.57 \pm 0.00 \times 10^{-1}$	NA	NA	$\mathbf{1.57 \pm 0.00 \times 10^{-1}}$
Texture	$\mathbf{4.34 \pm 2.67 \times 10^{-3}}$	$\mathbf{2.54 \pm 3.41 \times 10^{-4}}$	$\mathbf{4.58 \pm 1.31 \times 10^{-2}}$	$9.08 \pm 3.02 \times 10^{-2}$	$\mathbf{2.95 \pm 5.11 \times 10^{-5}}$
Satimage	$\mathbf{3.93 \pm 5.01 \times 10^{-6}}$	$\mathbf{1.66 \pm 2.83 \times 10^{-14}}$	$\mathbf{1.19 \pm 2.06 \times 10^{-6}}$	$\mathbf{4.14 \pm 6.05 \times 10^{-6}}$	$\mathbf{7.23 \pm 1.18 \times 10^{-8}}$
Breast Cancer	$3.17 \pm 0.00 \times 10^{-1}$	$1.63 \pm 1.35 \times 10^{-1}$	$3.17 \pm 0.00 \times 10^{-1}$	NA	$\mathbf{1.57 \pm 0.00 \times 10^{-1}}$
Crab Gender	$4.91 \pm 4.47 \times 10^{-1}$	$5.44 \pm 0.39 \times 10^{-1}$	$4.91 \pm 4.47 \times 10^{-1}$	$3.99 \pm 1.42 \times 10^{-1}$	$\mathbf{2.14 \pm 1.74 \times 10^{-1}}$
Digits (USPS)	$\mathbf{3.39 \pm 0.00 \times 10^{-2}}$	$7.60 \pm 0.00 \times 10^{-3}$	$\mathbf{3.39 \pm 0.00 \times 10^{-2}}$	$\mathbf{3.39 \pm 0.00 \times 10^{-2}}$	$\mathbf{2.80 \pm 0.00 \times 10^{-13}}$
CEDAR	$\mathbf{3.78 \pm 0.67 \times 10^{-2}}$	$\mathbf{1.60 \pm 2.20 \times 10^{-3}}$	$\mathbf{4.55 \pm 1.03 \times 10^{-2}}$	$\mathbf{4.22 \pm 1.44 \times 10^{-2}}$	$\mathbf{1.29 \pm 2.24 \times 10^{-10}}$
Arabic word	$\mathbf{1.31 \pm 2.28 \times 10^{-5}}$	$\mathbf{< 10^{-16}}$	$\mathbf{3.37 \pm 5.48 \times 10^{-5}}$	$\mathbf{1.49 \pm 2.23 \times 10^{-4}}$	$\mathbf{< 10^{-16}}$

The obtained results show that the proposed DWA is significantly different from the other combination rules, except for Iris, Breast cancer and Crab gender datasets. Indeed, the same performance are achieved in some cases assigned by not a number (NA) and small difference in other cases for which *p*-values exceed 0.05.

Furthermore, table 2.6 shows the ranking of various used combination rules. We can notice that the trained rules are more appropriate than the fixed ones. Moreover, the proposed DWA rule seems to be the best combiner since it achieves more accurate and stable results. On the other hand, the overall ranking shows that the Average is the best rule belongs to the fixed group, which confirms the findings from other studies (Kittler et al., 1998).

It is clear when the number of combined classifiers increases, the computational cost is also increased. However, since the OCC takes into consideration only the target class, the computational cost is less in terms of training time and memory space against some multi-class implementations such as OAO and OAA used with SVM classifiers. This demonstrates the main advantages of using OCC to solve the multi-class implementation.

Table 2.6: Ranking of the combination rules according to the used datasets.

Dataset	Combination rule					
	Average	Max	Prod	DT	SWA	DWA
Phoneme	2	5	2	6	4	**1**
Iris	**1**	6	**1**	5	**1**	**1**
Texture	3	5	4	6	2	**1**
Satimage	4	6	4	2	3	**1**
Breast Cancer	**3**	5	**3**	6	**1**	**1**
Crab Gender	5	2	5	3	4	**1**
Digits (USPS)	2	5	2	6	2	**1**
CEDAR	3	5	4	6	2	**1**
Arabic word	4	5	3	6	2	**1**
Overall	3.00±1.22	4.88±1.16	3.11±1.26	5.11±1.53	2.23±1.11	**1.00±0.00**

2.3.3.2. Influence of reduced OCCs

In order to study the stability for reduced classifiers from OCC ensemble, we propose to evaluate the performance of the MCS when only two OCCs are combined through the DWA rule. Table 8 reports the obtained results for various applications. We can observe that for Breast cancer, USPS and CEDAR datasets combining only two classifiers is more suitable than combing the five classifiers.

Figure 2.2 shows rates of MCS success deduced from table 2.7 (10 MCS and 9 datasets lead to 90 configurations). The objective of this comparison is to show the percentage of success when the MCS is better, equal or worse than the best single classifier. As we can see, the success rate of the MCS is more than 65.55% (59 times out of 90 comparisons), which proves the effective use of the proposed MCS for solving the multi-class classification problem.

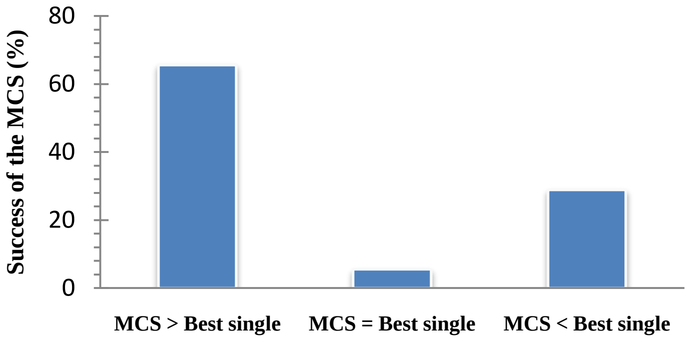


Figure 2.2: Rates of MCS success.

Table 2.7: MCS accuracy (%) for different OCC ensemble members.

OCC ensemble members	Phoneme	Iris	Texture	Satimage	Breast Cancer	Crab Gender	Digits (USPS)	CEDAR	Arabic word
SVM-K-Means	71.56±3.46	96.07±1.96	98.31±0.16	88.26±0.94	96.72±1.77	87.68±5.84	93.37±0.00	86.22±0.81	81.15±0.11
SVM-PCA	74.99±1.15	97.38±1.13	98.94±0.13	84.58±0.52	**97.00±1.48**	93.10±3.34	93.97±0.00	89.01±1.21	**86.43±0.56**
SVM-k-Center	74.91±1.82	96.73±2.99	98.73±0.23	**89.47±0.45**	96.58±1.86	90.61±2.15	93.17±0.00	88.19±1.17	80.66±**0.40**
SVM-NN	78.48±0.78	96.73±2.99	98.93±0.11	86.61±0.35	96.58±1.48	93.10±3.34	93.57±0.00	88.49±.125	**84.21±0.15**
PCA-K-Means	75.98±2.16	**98.03±1.96**	98.74±0.05	85.15±0.73	93.59±1.53	**93.12±4.70**	93.82±0.00	89.36±1.11	**85.15±0.43**
PCA-k-Center	75.56±1.01	96.73±2.99	**99.29±0.19**	87.75±0.56	93.86±1.95	93.10±3.34	**94.07±0.00**	89.07±1.07	**85.48±0.25**
PCA-NN	**79.17±1.17**	97.38±2.26	98.34±0.25	69.77±1.05	91.31±1.88	93.10±3.34	**94.07±0.00**	**89.66±0.87**	**86.45**±0.47
K-Means-k-Center	74.19±2.37	96.07±3.92	98.73±0.23	88.30±0.76	96.29±2.16	90.61±2.15	94.02±0.00	88.84±0.87	81.10±0.42
K-Means-NN	77.71±0.56	95.42±2.26	98.38±0.03	85.00±1.15	95.86±1.74	87.62±0.78	92.42±0.00	86.55±0.62	**81.15±0.11**
k-Center-NN	78.11±0.65	95.42±2.26	98.80±0.39	88.23±0.96	94.43±1.88	92.11±2.98	92.57±0.00	88.89±0.88	80.66±0.40
SVM	74.65±0.95	95.42±4.93	98.22±0.22	83.69±1.13	**95.16±2.76**	90.64±4.16	93.32±0.00	84.80±0.42	80.17±1.86
PCA	74.30±1.36	95.41±1.33	98.34±0.25	69.77±1.05	88.75±2.85	**93.10±3.34**	**93.97±0.00**	**87.87±0.60**	**83.34±0.78**
K-Means	70.66±3.72	**96.73±1.13**	96.37±0.34	85.26±0.99	95.01±1.78	86.70±6.64	92.27±0.00	86.75±1.70	80.95±0.31
k-Center	72.16±2.39	94.77±1.13	96.73±0.69	**87.28±0.85**	85.43+1.95	88.65±4.39	91.08±0.00	85.43±1.95	80.36±1.36
NN	**78.11±0.48**	95.42±2.26	**98.80±0.16**	84.47±0.79	91.31±0.98	**93.10±3.34**	92.47±0.00	86.01±0.61	82.43±0.86

2.3.3.3. Experimental proof of using the DWA

In order to prove the effective use of the DWA for combining an ensemble of multiple OCCs, figure 2.3 depicts an illustrative diagram of OCC ensemble showing relationship between three best classifiers deduced from the Phoneme dataset. Each region of the Venn diagram corresponds to a set of test samples recognized by a specific set of classifiers as reported in table 2.8. We can notice that each set of the used classifiers recognizes a specific rate of test samples. For instance, both OC-SVM and NN recognize 8.13% from the entire testing samples. Thus, attributing fixed importance for all classifiers cannot fit all testing samples. This drawback can be avoided when each classifier has its own contribution for each test sample through using the DWA. As explained above, when the classifier responds closely to its maximum response relatively to others, it is considered more certain to produce correct classification.

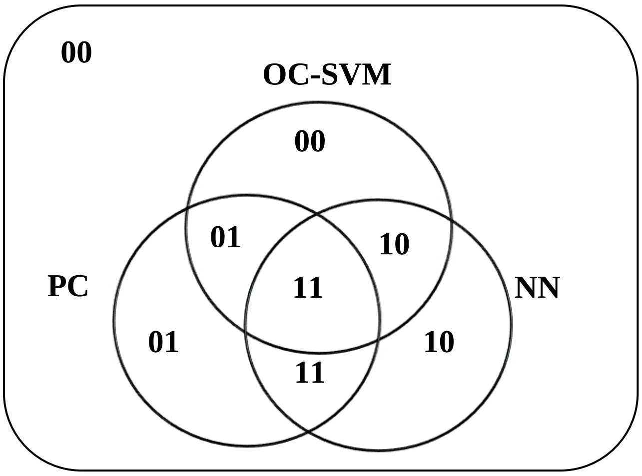
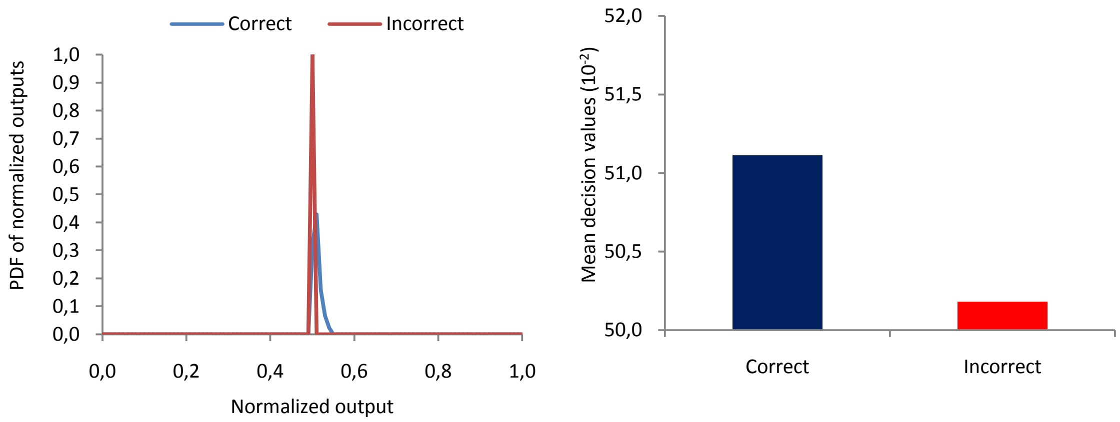


Figure 2.3: Venn diagram of OCC ensemble.

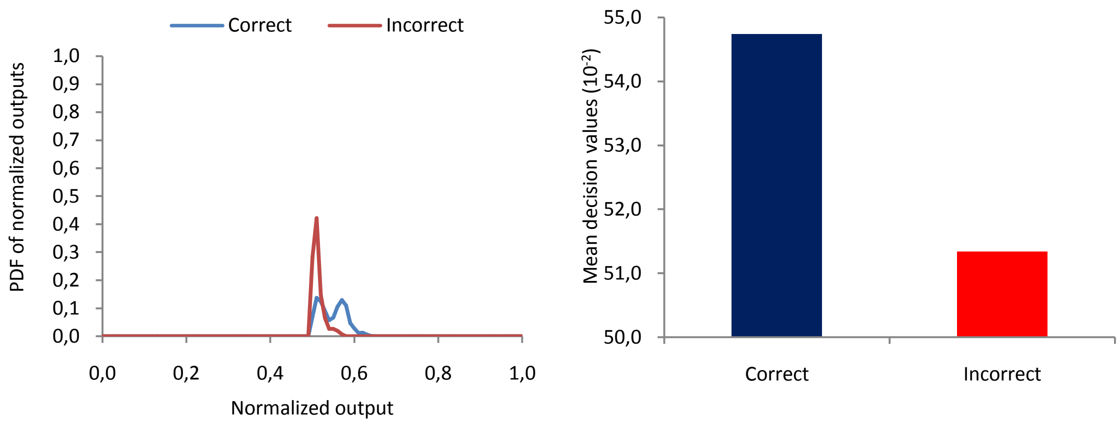
Table 2.8: Rate of test samples recognized by a classifier (√) and not by the other (×).

Code	SVM	PCA	NN	Rate (%)
001	√	×	×	10.17 ± 0.94
010	×	√	×	01.33 ± 0.34
100	×	×	√	01.51 ± 0.63
011	√	√	×	04.29 ± 0.56
101	√	×	√	08.13 ± 1.19
110	×	√	√	03.77 ± 0.49
111	√	√	√	64.90 ± 1.36

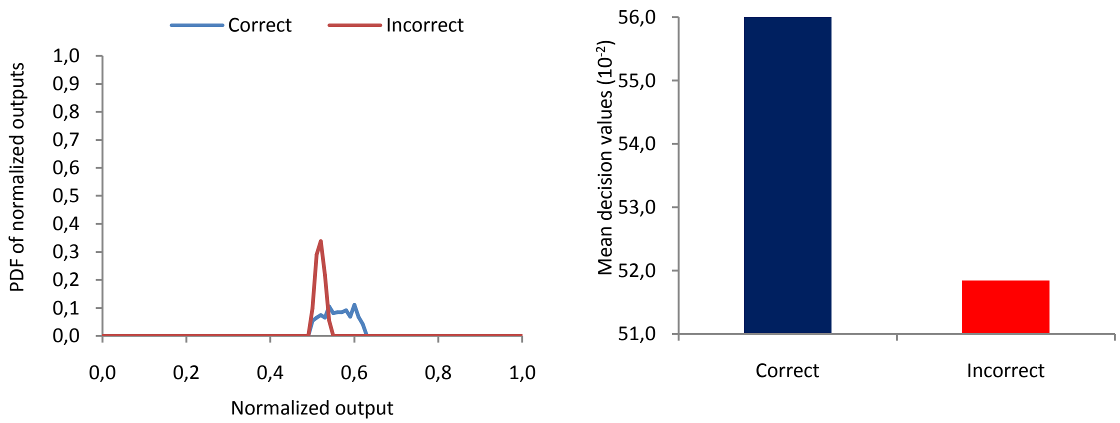
This fact is illustrated in figure 2.4 for which we present the Probability Density Functions (PDF) and mean values of classifier normalized outputs for both correct and incorrect



(a)



(b)



(c)

Figure 2.4: PDF and mean values for correct and incorrect classification (a) OC-SVM, (b) PCA and (c) NN.

classifications of Phoneme dataset. PDF and mean decision values show that outputs of correct classification are higher than the ones of incorrect classification. Therefore, the correct classification outputs are closer to their maximum responses, which allow detecting the more adapted classifier from the ensemble for each test sample via the proposed DWA rule.

To illustrate this point, we consider a test sample belonging to the first class of Phoneme dataset assigned correctly only by the PCA and OC-SVM. Table 2.9 reports the classifiers outputs, their corresponding posteriori probabilities, AERs, the maximum response values, static and dynamic weights generated through applying Eqs 2.3 and 2.5, respectively. On the other hand, Eq. 2.2 illustrates values of the decision templates computed for both classes.

As we can see, the k-Center achieves the lowest AER. Consequently, it generates the highest weight. However, this weight value is the same for all testing samples. When using the DWA combination rule, the maximum output value of OC-SVM is close to its maximum response which generates highest weight. In contrast, the maximum output value of the k-Center and K-Means are far from their maximum responses which lead to generate lower weights allowing to decrease their influence.

Table 2.9: Classifier outputs, posteriori probabilities, AER, maximum output values, static and dynamic weights for a test sample.

Classifier	op_i^1	op_i^2	$P_i(c_1/x)$	$P_i(c_2/x)$	r_i	op_i^{max}	w_i	$w_i(x)$
PCA	-0.2827	-0.0987	0.4792	0.5208	27.87	-0.0002	0.0359	0.2074
K-Means	-0.1620	-0.2198	0.5065	0.4935	28.85	-0.0003	0.0347	0.1947
SVM	0.0014	0.0018	0.4999	0.5001	26.70	0.0026	0.0375	0.2287
k-Center	-0.4024	-0.4689	0.5065	0.4935	26.31	-0.0003	0.0380	0.1531
1NN	-0.0603	-0.1273	0.5080	0.4920	27.80	-0.0032	0.0360	0.2162

$$DT_1 \begin{bmatrix} 0.5233 & 0.4767 \\ 0.5113 & 0.4887 \\ 0.5001 & 0.4999 \\ 0.5105 & 0.4895 \\ 0.5256 & 0.4744 \end{bmatrix} \text{ and } DT_2 \begin{bmatrix} 0.4879 & 0.5121 \\ 0.4717 & 0.5283 \\ 0.4999 & 0.5001 \\ 0.4983 & 0.5017 \\ 0.4883 & 0.5117 \end{bmatrix} \tag{2.7}$$

Table 2.10 reports the output values according to the used combination rule and its assigned class. We clearly can note that DWA is only the weighting method that achieves a correct classification. This example proves the efficiency of the DWA against the SWA rule and fixed rules. Thus, it seems that it is not easy to combine different types of OCCs~~.~~ Indeed, the behavior of the ensemble could not be learned by applying the DT rule. In summary, we clearly can conclude that linear combiners are more appropriate to deal with combining different types of OCC for the multi-class implementation.

Table 2.10: Output responses provided by different combination rules.

OCCs	Average	Product	Max	DT	SWA	DWA
Ensemble **1**	0.5001	0.3113	0.5079	0.9959	0.5001	0.4997
Ensemble **2**	0.4999	0.3112	0.5208	0.9957	0.4999	0.5003
Class label	1	1	**2**	1	1	**2**

2.3.3.4. Stability study of combination rules

In this section, we try to study the stability of various combination rules for an increased number of classes. Thus, we use CEDAR dataset via calculating the mean recognition rate by adding progressively new classes each time from 2 to 55. Figure 2.5 depicts the mean recognition rate versus the number of classes achieved by the various combination rules. Firstly, we can observe that for few classes, the most combination rules provide similar performances. However, when adding progressively new classes, we can see that the proposed DWA is the most stable.

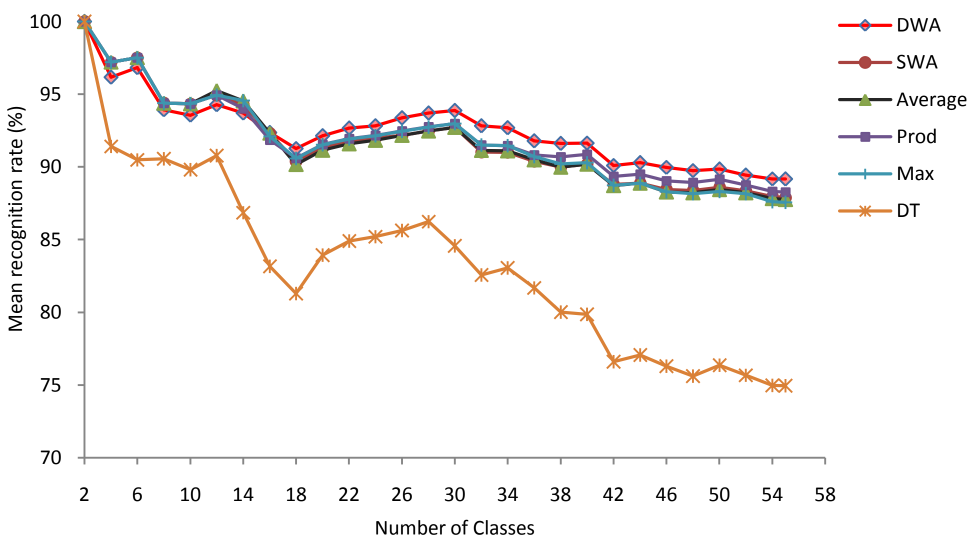


Figure 2.5: Stability study of combination rules.

Furthermore, Average, Prod, Max and SWA combination rules are also stable but provide less accuracy comparatively to the DWA rule. We also can notice that Average, Prod and SWA provide the same performance in terms of accuracy. A surprising result is the high instability and the low accuracy obtained by the DT, which demonstrates its sensitivity for an increased number of classes.

2.3.4. Comparative analysis

The main objective of using OCC for multi-class implementation is to achieve an open system. Hence, we aim to propose more accurate and robust open multi-class pattern classification based on DWA combination. Therefore, a comparative study is performed against a standard open classifier based on the K-Nearest Neighbor (K-NN) classifier (Duda et al., 2001).

Table 2.11 reports the obtained results for the K-NN classifier with different K values against the proposed MCS system. Roughly speaking, we clearly notice that the MCS achieves better results specifically for dataset having high number of classes for instance CEDAR and Arabic word datasets.

Table 2.11: Recognition accuracy (%) provided by the proposed MCS comparatively to K-NN classifier.

Classifier	Phoneme	Iris	Texture	Satimage	Breast Cancer	Crab Gender	Digits (USPS)	CEDAR	Arabic word
3-NN	78.07±0.94	96.80±1.07	97.90±0.21	87.95±0.22	96.45±1.36	81.86±8.49	87.48±0.00	77.98±0.99	67.19±0.45
7-NN	76.87±0.48	96.76±2.29	97.72±0.14	88.05±0.51	96.45±0.64	78.92±5.16	88.45±0.00	80.13±1.59	73.11±0.28
Our system	**79.52±0.45**	**98.69±1.13**	**99.49±0.45**	**89.55 ±0.65**	**96.86±2.20**	**93.61±4.21**	**94.02±0.00**	**89.17±0.59**	**85.37±0.34**

2.4. Conclusion

This chapter aims to study the combination of different types of OCC for solving the multi-class classification problem by means of a new combination strategy. Comparison has been performed by combining diverse OCCs *via* linear fixed and trained rules, in addition, to the well known DT strategy. Experimental results conducted on several real-world datasets show that combining diverse OCCs is defiantly worth it, since the MCS achieves better results than the best individual classifier. For the suitable combination strategy, we clearly can conclude that linear combiners are more appropriate to deal with combining different types of OCC for the multi-class implementation. Thus, a new DWA combination is proposed leading to more improvement of the linear combination and consequently the performance of the achieved MCS.

Besides, comparison of the proposed open classification system against a standard open classifier such as K-NN, shows the superiority of the proposed system specifically when the number of classes is high.

It is obvious that combining all classifiers is not necessary for all classes and datasets. Hence, the extension of this work consists to select for each class the most suitable OCCs in order to achieve a robust MCS. Furthermore, dynamic weighting is also a promising way to be explored for classifier aggregating.

Developing an effective combination rule is a standard way to perform improved OCC ensembles. However, in the next chapter of this thesis we present another way related to AANN ensembles improvement, which is based on an optimized selection of training samples.

Chapter 3

Optimized Selection of Training Samples for Improved One-Class Neural Networks Ensembles

Abstract

The one class Auto Associative Neural Network (AANN) has been investigated for solving various problems. Nevertheless, it is sensitive to the presence of outliers in the training set, which is known problem for one-class classifiers. For this, a straightforward approach consists of partitioning the training set into subsets leading to reduce the effect of outliers through a combination strategy. However, for the AANN even with ensemble method the effect of outliers is still maintaining. Thus, we propose an improved AANN ensembles based on a selection algorithm for selecting the appropriate training samples to the AANN, which allows enhancing the AANN combination and classification robustness. Experimental results conducted on several real-world datasets prove the effective use of the proposed algorithm to construct more robust AANN ensembles.

3.1. Introduction

Neural network is one among the most useful OCCs, which is usually referred to as auto-encoder and also as AANN. The AANN has been used for different applications. For instance, Cheedella et al., (2002) and Kishore et al., (2001) used the AANN for on-line text-independent speaker verification, Palanivel et al., (2003) used AANN for real time face authentication. In other work, Manevitz and Yousef, (2007) used the AANN for automated document retrieval and classification. In their application, the AANN is trained for filtering the documents under different conditions. In their application the AANN classifier proved its effectiveness to achieve better results than the Nearest Neighbor, Naive-Bayes, Distance-based Probability and one-class SVM algorithms.

Recently, the AANN has been extended to the multi-class classification problem for classifying cognitive states of brain activity (Boehm, et al., 2011). A genetic algorithm has been used for feature selection in order to enhance the recognition performances. AANN for multi-class classification has been also used by Leena et al., (2004) for language identification to distinguish between four Indian languages.

The most used AANN architecture is based on three layers (input, hidden and output layer). Usually, the AANN is used to compress the input data to less dimensions (for feature extraction), and subsequently to decompress these data back to original dimension in order to test the reconstruction ability. This classifier relies on training to reproduce the training dataset from the inputs to its outputs through adjusting parameters till minimizing the reconstruction error.

However, the main difficulty of using the AANN, is its considerable sensitivity to the presence of outliers or noisy data contained into the training set. Subsequently, the model induced by AANN may suffer from poor consistency when the training set includes abnormal training samples. Thus, Shieh and Kamm, (2009) proposed an OCC ensemble for combining OC-SVM using bagging method to reduce the effect of outliers. Hence, bagging method is proposed as an alternative solution for reducing the noise influence. However, for the AANN even with ensemble method the effect of outliers is still maintaining. Therefore, we propose an improved AANN ensembles (Hadjadji and Chibani, 2015) based on a selection algorithm for selecting the appropriate training samples (Hadjadji and Chibani, 2014b) to the combined AANN which allows enhancing the combination and classification robustness.

This chapter is organized as follows. Section 2 presents the algorithm used for selecting the most representative samples from the entire training set. Experimental results conducted on several real-world benchmarks are presented in section 4. Finally, the conclusion is provided in the last section.

3.2. Selection of Training Samples for AANN Ensemble enhancement

Several ways are possible for selecting the pertinent samples in order to reduce the outliers or noisy data. The most known method is based on support vectors developed by (Scholkopf et al., 2001) for OC-SVM classifier. In this work, we propose to investigate the AANN for detecting the outliers or noisy samples, then omitting them from the training set. This may lead to construct a robust representative model. The assumption relies on two main observations. Firstly, when the AANN learns from the entire training samples, it learns the internal structure of the consistence training samples. Secondly, the outliers or noisy samples are defined as samples that have unlike structure than the consistence ones. Therefore, the reconstruction error of these samples is higher than that for the consistence ones. For instance, if the training set contains 10% of outliers, according to our assumption, those represent 10% of the highest reconstructed error among the entire training samples. This is can be explained mathematically through the following equations (Hadjadji and Chibani, 2014b):

Denote S_{pr} and S_{out} the pertinent training sample set and the outlier set, which satisfy:

$$S = S_{pr} \cup S_{out} \tag{3.1}$$

The set fraction S_{out} containing N_{out} outliers is defined as a portion ν selected from the entire set Scontaining n samples.

$$S_{out} \subset S, N_{out} = \nu \tag{3.2}$$

Such that, $$0 < \nu < 1 \tag{3.3}$$

Consequently, the set fraction S_{pr} of the pertinent samples that contains N_{pr} samples is defined as follows:

$$S_{pr} \subset S, N_{pr} = (1-\nu)n \tag{3.4}$$

Denoting f_{ent} the representative model of AANN which is trained on the entire training set, the reconstruction error of each training sample x_i is then defined as:

$$Er(x_i) = f_{ent}(x_i) - x_i \tag{3.5}$$

We define an ordered set S_{ord} that contains the training samples which are ordered according to their reconstruction error from the minimum to the maximum value:

$$S_{ord} = \{\hat{x}_1, \dots, \hat{x}_n\} \tag{3.6}$$

Consequently, training samples are ordered according to their consistence value, such that $\hat{x}_1$ and $\hat{x}_n$ represent the most and the least consistence samples, respectively. Hence, the set of the most pertinent sample S_{pr} represents the first $(1-\nu)$ elements from the ordered set S_{ord} as:

$$S_{pr} = \{\hat{x}_1, \dots, \hat{x}_{(1-\nu)N}\} \tag{3.7}$$

Therefore, the optimal representation model is obtained through the following steps:

Step 1: Train the AANN on the entire training samples S to find the initial model f_{ent} .

Step 2: Order the training samples according the reconstruction error from the minimum to the maximum value.

Step 3: Select the pertinent training sample set S_{pr} according equation 3.7.

Step 4: Retrain the AANN on the pertinent training sample set S_{pr} to generate the optimal representative model f_{pr} .

The extension of the AANN to the multi-class classification is based on training each class on its respective AANN for a defined set of classes $C = \{c_1, \dots, c_m\}$, where m defines the number of classes. A test sample is assigned to the corresponding AANN when the best reconstruction is correctly achieved, (i.e. having the least reconstruction error Er). The class label $y(x)$ of a test sample x is defined mathematically as follows:

$$y(x) = \underset{j=1}{\overset{m}{\arg min}} \left(Er_j(x) \right) \tag{3.8}$$

In ordered to incorporate the selection algorithm with the ensemble method we generate randomly different datasets from the training set which allows generating various AANN for each class. The selection algorithm is applied for each AANN and finely, combination rules are used for aggregating the normalized outputs obtained from Selected Training Samples Auto Associative Neural Networks (STS-AANN) classifiers as it is presented in figure 3.1.

Note that, the out normalization method is the used softmax method which is explained in chapter 2 and 3. Besides, in this work we are not concerned about the combination method, hence, only fixed rules are used for evaluating the proposed approach.

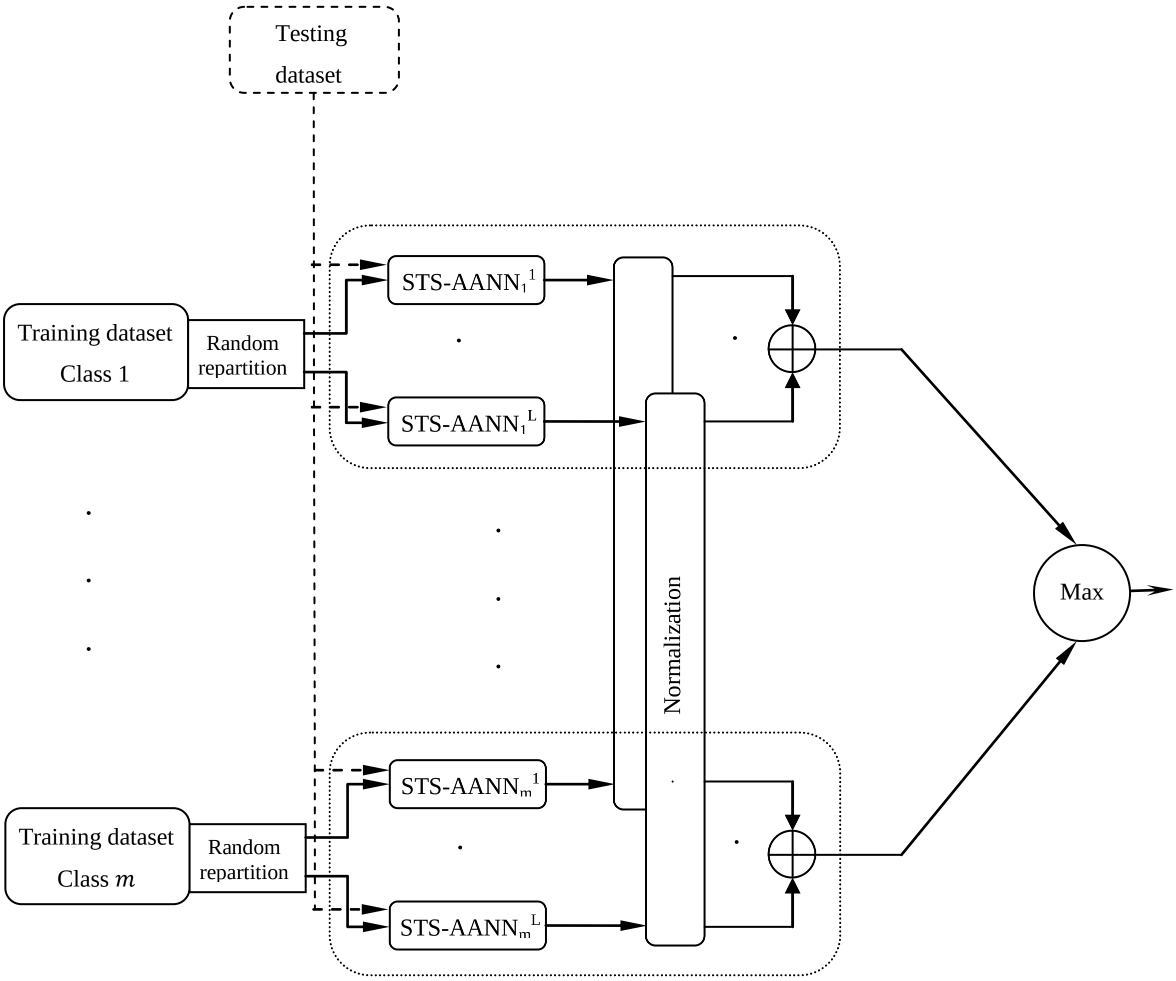


Figure 3.1: The overall architecture of STS-AANN ensemble.

3.3. Experimental Results

The proposed training samples selection algorithm on ensemble of AANN is evaluated using on several real-world benchmarks for solving the bi-class and multi-class classification problem which are reported in chapter 2. Similar to chapter 2 three-fold cross validation is performed for all datasets.

3.3.1. STS-AANN ensemble design and evaluation:

In order to create the ensemble of classifier we split up randomly the training set into three different sets for each class, therefore each class is represented by three classifiers trained on different samples (Hadjadji and Chibani, 2015).

For training the classifier, different parameters should be tuned. Firstly, the number of epochs is fixed at 200, which seems widely enough for the used datasets. Another parameter that should be carefully tuned to produce a correct representative model is the number of nodes in the hidden layer. Hence, each classifier is trained on the target class by varying the number of nodes. The optimal number is selected according to the best reproduction of the training dataset and consequently the least reconstruction error.

For evaluating the selection algorithm to enhance the combination we perform comparison between tow systems: ensemble of STS-AANN and AANN, which trained on the entire samples (with any selection).

Thus, the proposed selection algorithm is performed for selecting the best samples; the selection algorithm requires a parameter ν to be tuned between zero and one. This parameter is used for controlling the fraction or percentage of the outlier set. In this study, it is fixed as 10% from the entire training set. Consequently, ν is fixed at 0.1.

Performances of the classification are evaluated using Mean Recognition Rate (MRR). Table 3.2 reports MRR for both systems based on ensemble of AANN and STS-AANN, respectively.

As it can be seen from the Table 3.1, training the ensemble of AANN on the entire training set affects considerably the results on all the used datasets. Indeed, for a reduced number of samples, the MRR of the STS-AANN ensemble is higher comparatively to the AANN one, which is trained on the entire set. Besides, it is worth a while to note that, ensemble of STS-AANN outperforms the ensemble AANN whatever the used combination rule as it reported in Table 3.1. Therefore, the most important for generating an effective ensemble is not only partitioning the training set, but checking the consistence of the training set.

Table 3.1: AANN ensembles against STS-AANN with different combination rules

Combination rule	Average		Max		Min		Prod	
Dataset	AANN	STS-AANN	AANN	STS-AANN	AANN	STS-AANN	AANN	STS-AANN
Banana	87.65±3.57	**91.07±2.64**	89.60±5.88	**90.58±1.79**	89.60±5.88	**90.58±1.79**	89.60±5.88	**91.07±2.64**
Cancer	96.43±0.66	**96.57±1.14**	96.15±0.44	**96.29±1.32**	96.15±0.44	**96.29±1.32**	96.43±0.66	**96.57±1.14**
Iris	97.38±2.26	**98.03±1.96**	**98.03±1.96**	**98.03±0.00**	**98.03±1.96**	**98.03±1.96**	**98.03±1.96**	**98.03±1.96**
Phoneme	74.28±0.89	**76.97±2.39**	74.06±0.86	**76.91±2.65**	74.06±0.86	**76.91±2.65**	74.01±0.78	**76.39±2.13**
Crab	63.77±17.35	**66.35±1.74**	80.03±14.50	**85.59±4.81**	80.03±14.50	**85.59±4.81**	79.54±4.25	**86.08±5.05**
Satimage	78.37±3.62	**80.96±1.72**	76.13±3.92	**79.50±1.27**	79.10±3.28	**79.50±1.27**	78.49±3.61	**81.07±1.61**
Texture	95.46 ±0.52	**95.69±0.42**	95.26±0.51	**95.53±0.99**	93.84±1.23	**93.86±1.17**	95.40±0.51	**95.64±0.47**

In order to prove the effective use of the proposed approach, we present an example of Banana dataset. Figure 3.2 shows the training samples of Banana dataset, which are mapped in their reconstruction error space generated by the initial model f_{ent}. The red dots and stars represent the first and the second classes, respectively. REF 1 and REF 2 denote the reconstruction errors of feature 1 and 2, respectively.

In order to select the most representative samples, the proposed selection algorithm is applied separately on each class. The selected samples are showed by the blue circles for the first and second classes, respectively.

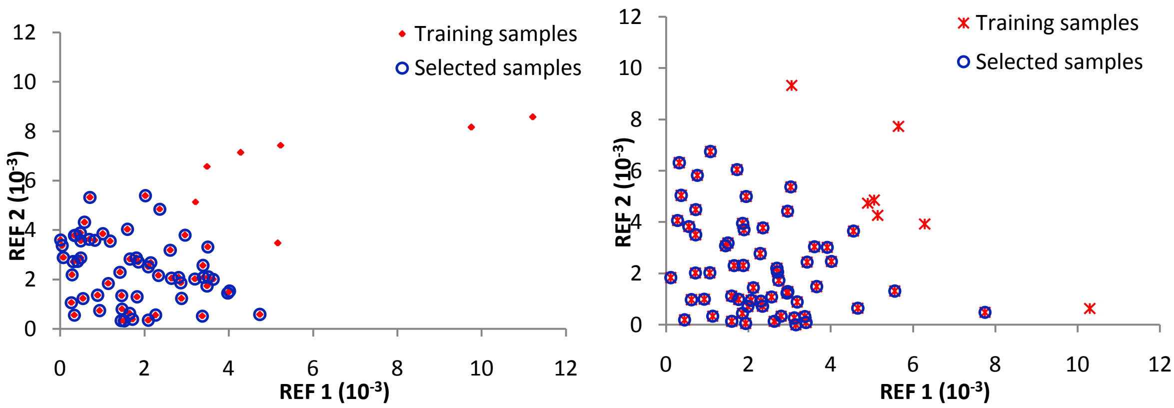


Figure 3.2: Mapping the selected and unselected training samples in the reconstruction error space

We clearly note that, the consistence training samples are near to the center and thus, they are well reproduced by the model, which learnt the internal structure of target class samples. On the other hand, the outliers, which are not selected, are distributed far from the center. Therefore, the reconstruction error of these samples is higher than that for the consistence ones, which means that they have unlike structure than the consistence ones. As a result, the model that learns the internal structure from the training samples offers an easy way for detecting the outliers or noisy samples.

Figure 3.3 shows the distribution of the training samples in their original space. As it can be seen, samples which are not similar to the majority are not selected by the algorithm. On the other hand, the outliers are located in the borders of the training sample distribution.

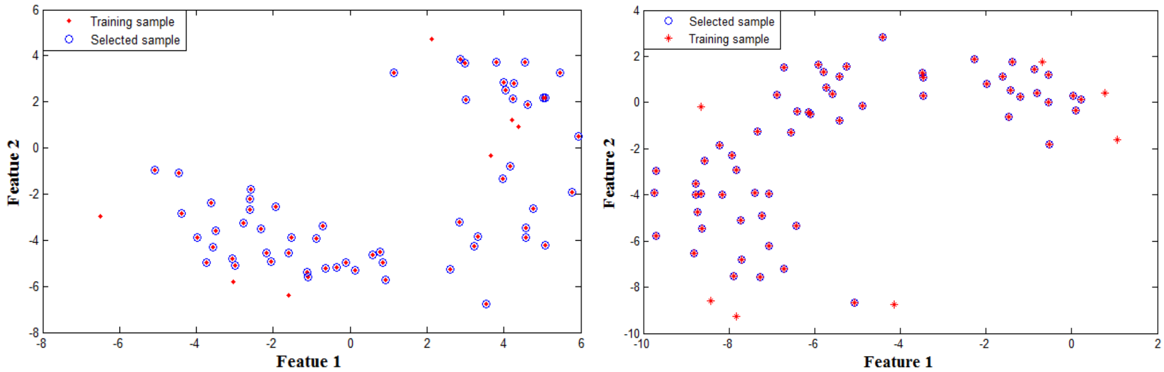
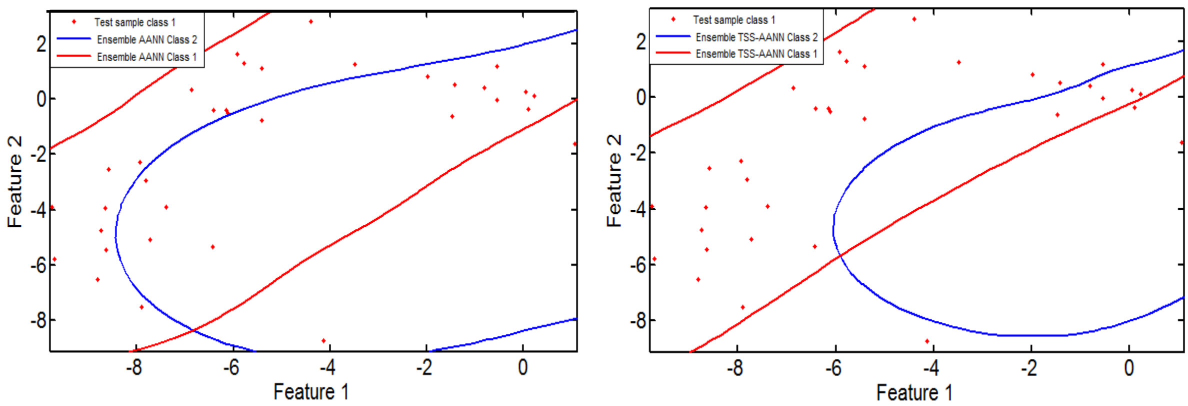


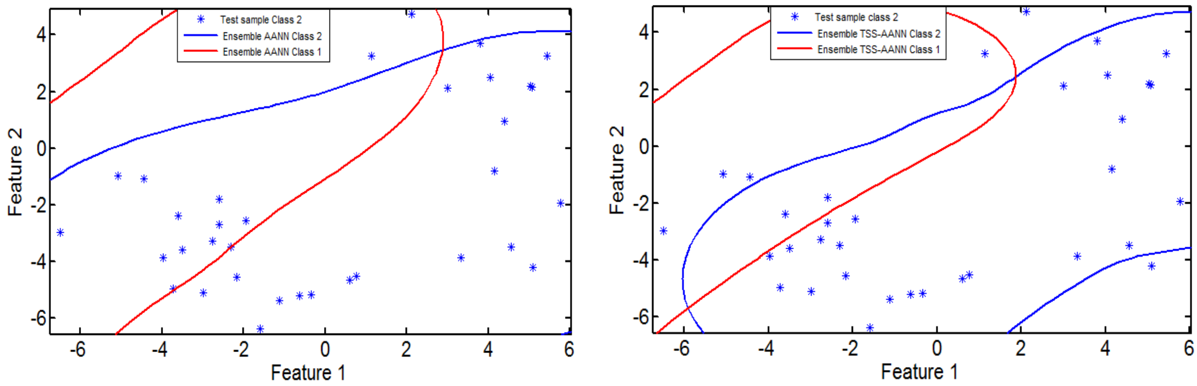
Figure 3.3: Mapping the selected and unselected training samples in the original space for both classes

Figure 3.4 shows the effect of training sample selection through mapping test samples on both ensembles AANN and STS-AANN. We can clearly see the effectiveness of the selection algorithm to reduce the overlap between the classes and therefore reducing miss classification.

Figure 3.4 also proves that using ensemble of AANN alone to reduce the effect of outliers is not enough, and their effect is still maintaining. This has guided us to joint AANN ensembles with a selection algorithm, which proves its effective use for more enhancement of the ensemble robustness.



(a)



(b)

Figure 3.4: Effect of training samples selection on the test samples on both ensembles for class 1 (a), class 2 (b) using mean combination rule

3.4. Conclusion

The objective of this chapter aims to propose an alternative method for ensembles improvement, which is always performed on the combination rule. Thus, the proposed method relies on an effective selection of training samples for performing an improved AANN ensemble. More precisely, the pertinent samples are selected according to their reconstruction error calculated between the input and its corresponding output generated by an initial model. Experimental results conducted on several real word datasets reveal that, the proposed selection algorithm is very easy to implement and allows improving effectively the classification accuracy whatever the combination rule.

In the continuation of the presented thesis, we will explore the hybrid OCC ensembles for handwritten recognition applications which are, Arabic handwritten word recognition, and writer identification form handwritten signature as well as handwritten text fragments.

Chapter 4

Hybrid One-Class Classifier Ensembles based-Fuzzy Integral for Open-Lexicon Handwritten Arabic Word Recognition

Abstract

So far, hidden Markov model, support vector machine and neural networks are the most used classifiers for Arabic word recognition, which provides a system with closed lexicon. In this chapter, the OCCs are explored in order to perform an Arabic word recognition system with an open lexicon. Generally, pattern recognition systems designed by a single system suffer from limitations such as the lack of uniqueness and non-universality. Thus, combining multiple systems becomes an attractive research topic for performance and robustness enhancement. Usually, fixed rules are the standard used combiners for the hybrid OCC ensembles. Thus, the present chapter aims to propose a combination scheme of OCCs based on fuzzy integral operators. Furthermore, an alternative framework is proposed to design a parameter-independent and open-lexicon handwritten Arabic word recognition system as well as a new density measure function. Experimental results conducted on Arabic handwritten dataset using different types of OCCs with large number of classes show the superiority of FI for hybrid OCC ensembles.

4.1. Introduction

Handwritten Arabic word recognition is an active research field due to its interesting use in different applications such as automatic sorting of postal mail, automatic bank check processing, bills processing, passport validation and recently, for historical document reading via text to speech applications, helping blinds to read and recognizing handwritten historical documents (Lawgali, 2015; Shatnawi, 2015; Alginahi, 2013; Likforman-Sulem and Sigelle, 2008). Arabic language is unlike the Latin languages, which is written from right to left. Also, it has its own diacritical marking such as dumma (’), hamza (ء), and the madda (~). Regarding character shapes, Arabic script has two main properties. On one hand, several letters share the same shape and differ only in the number and position of dots, such as “djim: ج”, “ha: ح” and “kha: خ”. On the other hand, some letters change their shape according to their situation at the beginning, the medial or the end in the word. For instance, the letter “Aïn”, can be written through four shapes that are: “ عـ, ـعـ , ـع , ع”, where the two last shapes are related to end positions which change if the word is fully connected or not.

So far, the Arabic word recognition is considered as one of the most challenging task in pattern recognition for its specific writing as well as its variability. In this context, the analytical and holistic approaches are the two possible ways for recognizing an Arabic word (Farah et al, 2005). The first one consists to segment a word image into sub-words or isolated characters, which are recognized through character recognition. Generally, the analytical recognition is useful when very large vocabulary is available since it is impossible to construct a specific classifier for each word. Hence, the Hidden Markov Models (HMM) are considered as the most probably used classifiers for solving very large vocabulary lexicon (Knerr et al, 1998). In contrast, the holistic approach is based on the global analysis where each word is considered as a single unit, which is appropriate for problems with large or medium vocabulary such as address postal recognition (Plötz and Fink, 2009). In this case, all kinds of classifiers can be used such as the binary Support Vector Machines (SVM) (Muñoz-Marí et al, 2009; Krawczyk and Filipczuk, 2014). The main advantage of this approach is related to its efficient capture of the co-articulation and variability effects contained into word images handled by the same classifier (Vinciarelli et al, 2002). Hence, the present work is focused on the use of the holistic approach for recognizing handwritten Arabic word.

So far, HMM (Jayech et al., 2016), SVM (Hmeidi et al., 2008), neural networks (AlKhateeb et al., 2011) and hybrid of SVM and conventional neural networks (Elleuch et al., 2016) are the most used classifiers for Arabic word recognition. However, the use of such type of classifier provides an Arabic word recognition system with closed lexicon, indeed, adding new word to the lexicon requires retraining all the system.

Nowadays, extended multi-class implementation to new classes is strongly required, for instance, in Arabic word recognition and handwritten writer identification. Nevertheless, the existing classifiers need to retrain the system again on all classes such as the One-Against-One (OAO) or One-Against-All (OAA) implementations based on the SVM classifiers.

In this chapter, the OCCs are explored in order to perform an open-lexicon Arabic word recognition system. Generally, pattern recognition systems designed by a single system suffer from limitations such as the lack of uniqueness and non-universality to the problem at hand (Kwak and Pedrycz, 2005). Thus, combining multiple systems by taking advantage of each individual and avoiding their weakness may lead in the improvement of classification accuracy. Therefore, combining classifiers trained on different features intending to investigate their complementary characteristics is definitely worth it. Indeed, the benefits of multiple classifiers based on different features for the same problem have been suitable for various fields of pattern recognition, including handwritten recognition (Chiang and Gaber, 1997), speech verification (Pham and Wagner, 2000) and other applications (Chiang, 1999).

Recently, it has been demonstrated that combining classifiers can also be effective for OCCs and therefore, the existing classifier combination strategies can also be applied to OCCs (Yeh et al., 2009). Since information regarding only one class is available, combining OCCs becomes more difficult.

Usually, the combination step for the hybrid OCC ensemble is performed through the use of simple combination rules such as fixed rules (Juszczak, R.P.W. Duin, 2002; Muñoz-Marí et al., 2007), ECOC and DT strategies (Krawczyk and Filipczuk, 2014; Krawczyk and Wozniak, 2014). However, fixed combiners cannot be useful to treat some difficult cases. Fixed rules are optimal for special cases for which the combined systems are similar in terms of performance and competence. Moreover, classifiers designed by various information sources are different from each other, because the members of the ensemble are built of diverse feature spaces (Kuncheva, 2004). Therefore, trained combiners are more suitable since a priori-knowledge about the ensemble is investigated in the test phase, which allows giving more importance to the more suitable classifier. Thus, the final decision is made by taking into account the competence of each member.

In this respect, a great effort has been done for proposing various combination methods and schemes including methods based on the Dezert-Smarandache Theory (DSmT). Indeed, (Abbas et al., 2013) proposed DSmT a new scheme based on OC-SVM ensemble trained on different feature sets using the DSmT for handwritten digit recognition. The DSmT shows its superiority in term of performance versus the sum rule. However, the proposed scheme violates the best advantage of using OCCs as multi-class system since the extension to new classes, achieves a closed system. Indeed, adding new classes requires updating all parameters and retrains the combination model.

Other alternative combination methods have been proposed based on fuzzy sets. In that case, FI and the associated fuzzy measures initially introduced by Sugeno have been reported to give excellent results for classifier aggregation. Its main advantage is related to measuring the strength not separately for each classifier alone but for all members. The ability of the fuzzy integral to enhance the results produced by multiple information sources has been proved in various application areas of pattern recognition (Kwak and Pedrycz, 2005; Chiang and Gaber, 1997; Pham and Wagner, 2000; Chiang, 1999).

For pattern classification usually parameters should be adjusted each time when a new class is added to the system. Besides, in FI the density measure representing the ensemble competence of each member is measured by its achieved performance. Consequently, all testing samples are represented by the same density measure values, which make this approach less efficient since each test sample has its suitable density measure. Therefore, more contribution should be given in respect to the more appropriate information source and each test sample.

To overcome these drawbacks, this chapter proposes to investigate an alternative framework for designing a parameter-independent open-lexicon Arabic word recognition system (Hadjadji et al., 2014c) as well as a new density measure function. This framework allows providing a dynamic measure for each test sample without the need to measure its performance *via* the training or validation dataset. Since OCCs have not been evaluated yet for large number of classes, results are carried out on large dataset for Arabic handwritten word recognition and different types of OCC, leading to have an extended view on the usefulness of proposed framework based on FI to the addressed problem.

The remaining of this chapter is organized as follows. Section 2 describes the mathematical formulation of the proposed hybrid OCC ensemble. In section 3, experimental results are conducted on various types of OCC for handwritten Arabic word recognition with large number of classes in order to prove the effective use of the proposed combination scheme. Finally, the conclusion and future work are provided in the last section.

4.2. Hybrid OCC ensemble based on fuzzy integral

The hybrid system is composed of different single OCC ensembles, where each one is dedicated to represent one-class from the set of classes. Consequently, each class is represented by a combination of different OCCs. Various combination rules are possible for achieving an enhanced hybrid OCC ensemble. The present chapter proposes to investigate the FI operators for combining multiple single OCC ensembles.

The hybrid OCC ensemble depicted in figure 4.1 is composed of m classes and L different information sources. Therefore, each class is represented by a single OCC ensemble which is composed of L OCCs trained on different information sources. Their normalized outputs are aggregated through a FI operator. Finally, the class label of the test pattern is assigned to the single OCC ensemble that achieves the maximum prediction.

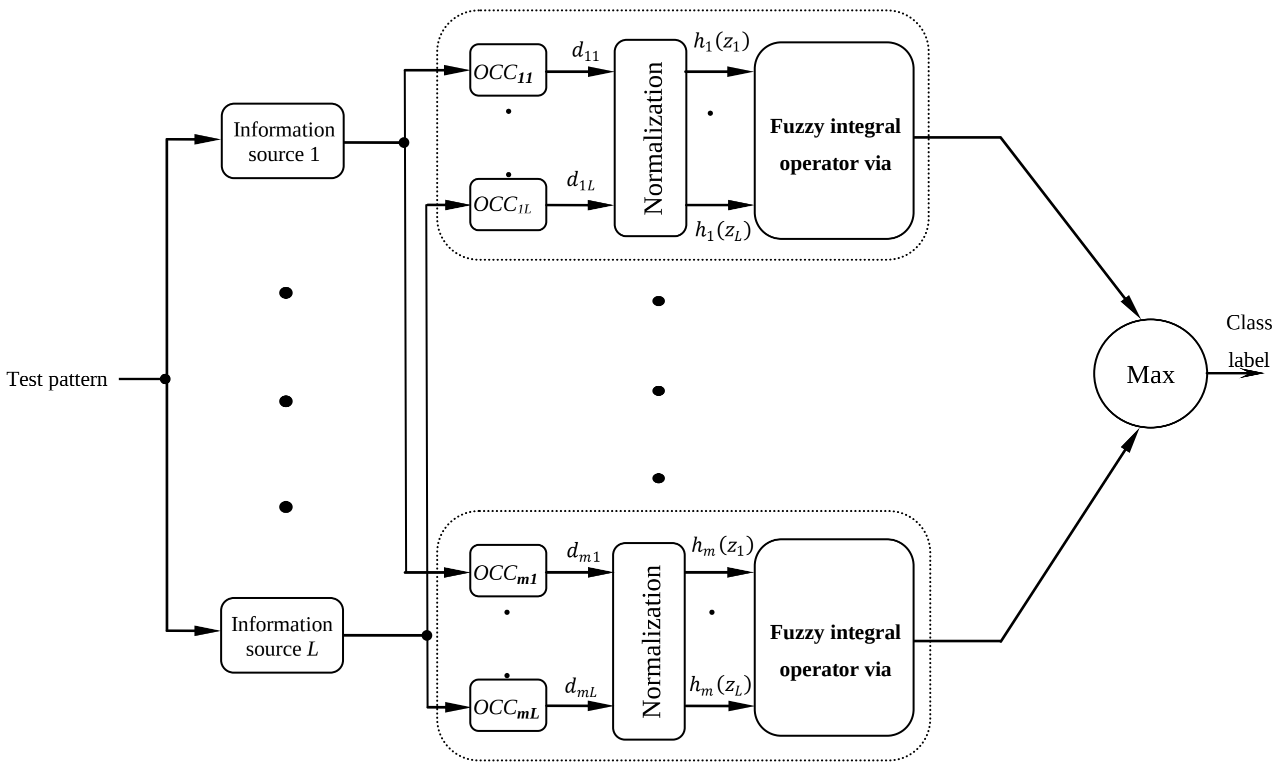


Figure 4.1: Hybrid OCC ensemble scheme based on Fuzzy integral.

Figure 4.1: Hybrid OCC ensemble scheme based on Fuzzy integral.

Let $\{D_i, i = 1,..m\}$ be the set of m single OCC ensembles and denote $D_i = \{d_{ij}, j = 1,..,L\}$ as the output vector composed of the output value d_{ij} provided by the OCC_{ij} trained on the i^{th} information source of the j^{th} class. The set of the output values can be represented in a matrix as follows:

$$D = \begin{pmatrix} D_1 \\ D_2 \\ . \\ . \\ . \\ D_m \end{pmatrix} = \begin{pmatrix} d_{11} & d_{12} & \cdots & d_{1L} \\ d_{21} & d_{22} & \cdots & d_{2L} \\ . & . & \cdots & . \\ . & . & \cdots & . \\ . & . & \cdots & . \\ d_{m1} & d_{m2} & \cdots & d_{mL} \end{pmatrix} \tag{4.1}$$

Several combination rules are possible to achieve the hybrid OCC ensemble, but all these rules need a unique interpretation of the outputs generated by the different classifiers for each test pattern x. Hence, the normalization of outputs for each classifier is required for performing correctly the combination. In this work, the simple exponential function is used for transforming the OCC output d_{ij} ranging from $]-\infty, 0]$ to $]0,\ 1]$ using the posteriori probability $P_i(c_j/x)$ as follows:

$$P_i(c_j/x) = exp\left(d_{ij}(x)\right) \tag{4.2}$$

The evidence is then expressed as the posteriori probability taking the following form:

$$h_j(z_i) = P_i(c_j/x) \tag{4.3}$$

The successful key of FI depends on the appropriate formulation of the density measure associated to each information source. Consequently, if the density measures are well formulated then the fuzzy measures can be correctly defined, in order to make a well aggregation of the fuzzy integral. Generally, the density measures represent the competence of the ensemble member measured by its accurate performance. The use of the performance deduced from the training datasets is not well representative since it requires a new validation dataset to get a better evaluation of each information source. Moreover, all testing samples are represented by the same values, which make this approach less efficient. Indeed, more contribution should be given in respect to the more appropriate information source and each test sample. To overcome this drawback, we propose an alternative approach in which the density measure is considered as the similarity degree or correspondence between the class model and that of the test pattern, i.e. the closer the similarity is, the greater the value of the density measure is. Therefore, a new dynamic density measure is proposed, which is defined for each test sample as:

$$g_{ij}(x) = exp\left(-\delta\left\|d_{ij}(x) - dm_{ij}\right\|^2\right) \tag{4.4}$$

Such that, $0 \leq g_{ij}(x) \leq 1$ and $0 < \delta \leq 1$.

The dm_{ij} is obtained by averaging values of the OCC_{ij} outputs using the training samples. δ is a positive value that has been introduced for calibrating more efficiently the density and therefore representing more accurately the contribution of each classifier and information source.

Let $y(x)$ be the class label of a test pattern x and one of the defined FI operators Sugeno (I_S), Choquet (I_C), S-AND (I_{S-AND}), C- AND (I_{C-AND}) and OR (I_{OR}), the aggregation of posteriori probabilities of the different information sources and their corresponding fuzzy measures to be combined is defined as:

$$y(x) = \underset{j=1}{\overset{m}{argmax}} \left(FIOP_j\left(h_j(z_i),\ g_j(A_i)\right)\right), 1 \leq i \leq L \tag{4.5}$$

$FIOP$ is one of the defined FI operator.

In summary, the multi-class classification using the FI is performed according to the Algorithm 4.1.

Algorithm 4.1:

Inputs: Image word I, classifier models OCC_j^i, average training outputs dm_{ij}

Output: Class label

1: for $i \leftarrow 1$ to m do /* i represents class variable */

2: for $j \leftarrow 1$ to L do /* j represents the feature generation method variable */

3: Calculate the word feature vector x of the image word I

4: Calculate the classifier output $d_{ij}(x)$

5: Compute the normalized output $P_i(c_j/x)$ and the evidence value $h_j(z_i)$ according to Eq. 4.2 and 4.3.

6: Calculate the density measure according to Eq. 4.4

7: end for

8: Determine the fuzzy measures $g_j(A_i)$, for $1 \leq i \leq L$ and $1 \leq \mathrm{j} \leq \mathrm{m}$ using Eqs. 1.24 and 1.25.

9: Perform FI operator to determine the combination score via the evidence and fuzzy measure

10: end for

11: Assign the unknown word to the word class that provides the maximum FI combination score according to Eq. 4.5.

4.3. Experimental results

4.3.1. Dataset description

In order to evaluate the proposed approach on large number of classes, the well known IFN/ENIT dataset (www.ifnenit.com) is selected containing more than 26400 images of Tunisian town names written in Arabic script. Words are written by 411 writers using different writing tools. The IFN/ENIT is composed of four folds, A, B, C and D. Usually, results are carried out using three folds for training and one for testing. The present work is evaluated using 300 classes each one is trained by considering more than 10 samples per class. All results are reported in terms of the classification accuracy expressed in percentage (%).

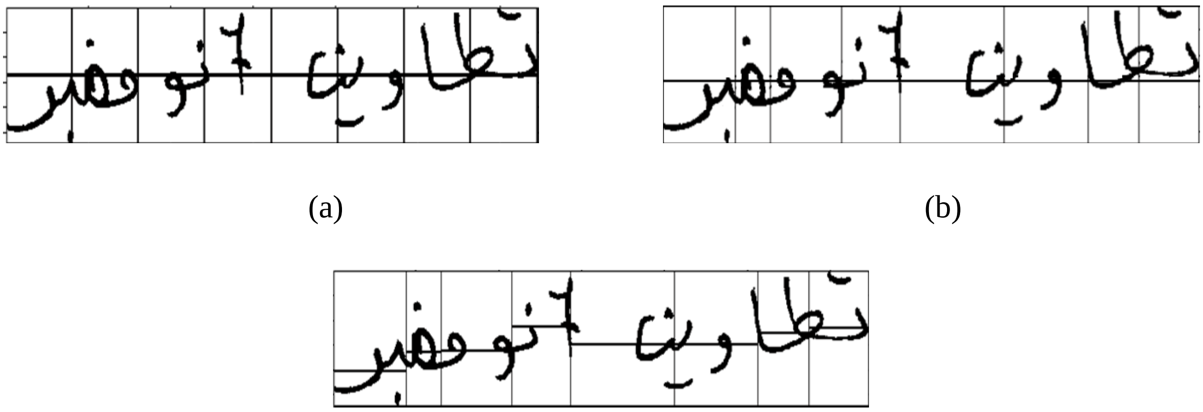
4.3.2. Information source generation

Various techniques have been proposed for generating features from the word image (Rath and Manmatha, 2003). In this work, the Curvelet transform (CT) is used for its enhanced directional capacity to characterize edges and singularities along curves that compose handwritten Arabic word (Candès et al., 2006). CT has been employed in various applications such as image denoising (Starck et al., 2002), face and facial expression recognition (Mandal and Wu, 2008; Saha, and Wu, 2010), compression (Majumdar, 2009), texture classification

(Arivazhaganet al., 2006), content based image retrieval (Sumana et al., 2008), character recognition (Majumdar, 2006; Kazemi et al., 2008). Recently the CT has been successfully used for offline handwritten signature retrieval (Shirdhonkar and Kokare, 2011) and verification (Guerbai et al., 2015).

Presently, CT is first performed on the word image *via* the wrapping technique at different scales and different orientations to generate curvelet coefficients. The resulting ones are used for computing the energies, which allow characterizing the importance of the curvature contained into the word image. In order to capture more efficiently the local information, CT is performed on different sections of the word image grid. Finally, the feature vector is achieved by concatenating all computed wedge energies for the defined image sections.

In order to perform the combination, different information sources should be generated from the word image (Abbas et al., 2013; Rath and Manmatha, 2003). Also, three different ways are used for partitioning an image into variety of sections (before applying the curvelet transform) namely equispaced, equimass and equimass adaptive grids, respectively. A uniform or equispaced grid (Favata and G. Srikantan, 1996) creates rectangular regions for sampling, where each one has the same size and shape. It is performed via placing the grid lines at equally spaced locations along the x-axis of the word image creating the vertical regions. Similarly, the horizontal regions are produced by placing grid lines at equally spaced locations down the y-axis. Conversely, an equimass grid creates different rectangular regions having the same number of black pixels, also known as the mass, of the word image (Favata and G. Srikantan, 1996). Consequently, each region is found by partitioning horizontally and vertically the word image using its mass histogram. Hence, the total mass between all adjacent points on either the x-axis or the y-axis are as close to equal as possible. Additionally, the equimass adaptive is used, which is a small modification of the equimass grid. It is based on the computation along the x-axis for each horizontal region and not for the entire image as in equimass grid. Figure 4.2 shows an example of partitioning the handwritten word image using three methods.



(a) (b)

(c)

Figure 4.2: Arabic handwritten word decomposition having the same grid size 2×8

(a) Equispaced grid, (b) equimass grid and (c) equimass adaptive grid.

4.3.3. Experimental design

In order to design the hybrid OCC ensemble parameter-independent and open-lexicon Arabic word recognition system, experiments are conducted into two steps: design step and evaluation step. During the design step, a number of classes is randomly selected from the whole dataset in order to deduce the optimal parameters for feature generation, training parameters for OCCs and the combination model. During the evaluation step, the remaining classes are used for evaluating the robustness of the proposed system taking the same parameters found during the design step. In other words, when a new class is added to the hybrid OCC ensemble for evaluation, the same parameters are used, as they have been tuned during the design step.

In this work, the hybrid OCC ensemble is composed of three OCCs for each class, each one receives its own feature vector according to the selected feature generation method. Hence, each OCC is separately trained on its own information source. In order to have a large view on the usefulness of the proposed architecture, results are carried out on different types of OCCs which are PCA, K-Means and NN (Tax, 2001). Therefore, different systems are built according to the used type of OCC. These classifiers are selected for their success in many applications and for the reduced numbers of parameters to be tuned during their training.

During the design step, the proposed open classification system is highly affected by the grid size and the classifier parameters. For the NN classifier, no parameters are required to be tuned unlike to other classifiers. For the K-Means, an only parameter should be carefully tuned corresponding to the number of cluster. The PCA also needs an only parameter corresponding to the number of eigenvectors. For finding the optimal parameters of the grid size and classifiers, 10 classes are selected randomly having more than 10 samples per class. To build the OCC models, the set of samples for each class is divided into two subsets namely Subset 1 and Subset 2, respectively. Subset 1 is used for training the OCCs while Subset 2 is used for finding the optimal parameters of the grid size and classifiers.

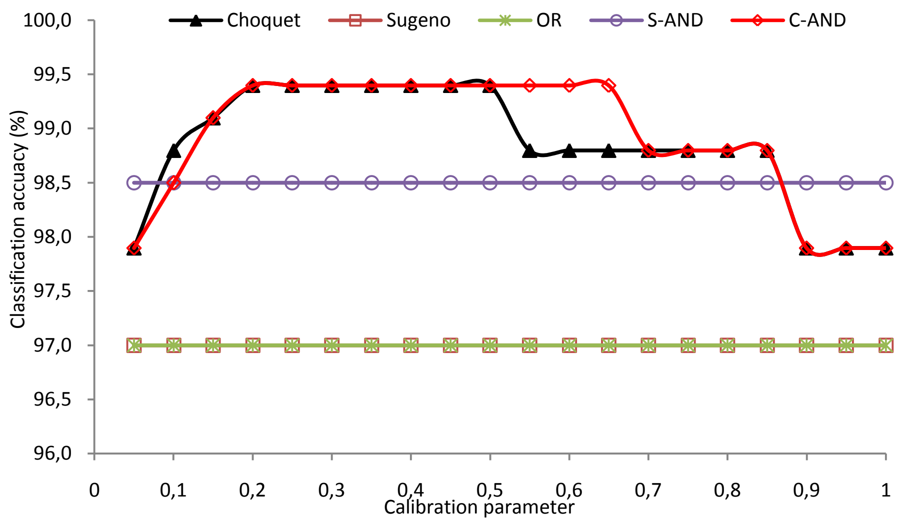
Results expressed in terms of classification accuracy for three information sources with different grid sizes obtained with the best classifier parameters are reported in table 4.1. As can be seen, the grid size affects significantly the classification accuracy whatever the used classifier. Indeed, when the grid size increases, the classification accuracy is also enhanced. Therefore, the suitable grid parameters are selected to achieve the best accuracy. In this case, the grid size 2×8 offers the best accuracy and optimal feature vector size for all classifiers and for three grid types. Therefore, this grid size is selected for all next experiments.

Table 4.1: Classification accuracy (%) of individual OCCs with various equispace (ES), equimass (EM) and equimass adaptive (EA) grid sizes.

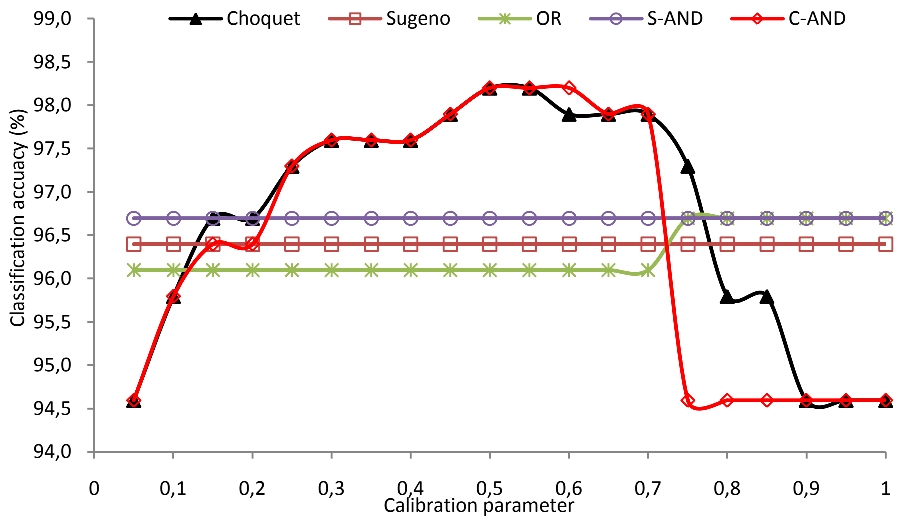
		PCA			K-means			NN		
Grid size	# Features	ES	EM	EA	ES	EM	EA	ES	EM	EA
2×2	192	80.78	84.98	86.78	81.98	81.38	82.23	74.17	74.17	74.40
2×4	384	91.89	93.39	93.99	89.18	91.29	91.89	81.38	84.38	85.70
2×6	576	94.29	95.19	95.79	92.49	92.49	95.19	84.08	86.78	89.18
2×8	768	**95.79**	**96.09**	**96.69**	**93.09**	**93.99**	**96.69**	**93.09**	**93.99**	**96.39**

Once designing the individual classifier, the combination is performed by means of the FI operators. As already reported, the use of FI requires a careful tuning of parameters, which are the calibration parameter (δ) related to the fuzzy densities according to Eq. 4.4, and the couple (α, β) related to OR, C-AND and S- AND operators, respectively. Similarly to individual systems, suitable parameter (δ, α, β) values are tuned using also the 10 classes. Results for the used operators with different values of the calibration parameter (δ) ranging from 0.1 to 1 with a step 0.1 are depicted in figure 4.3. Moreover, for OR, C-AND and S-AND operators, the presented results are obtained with best parameters of the couple (α, β)

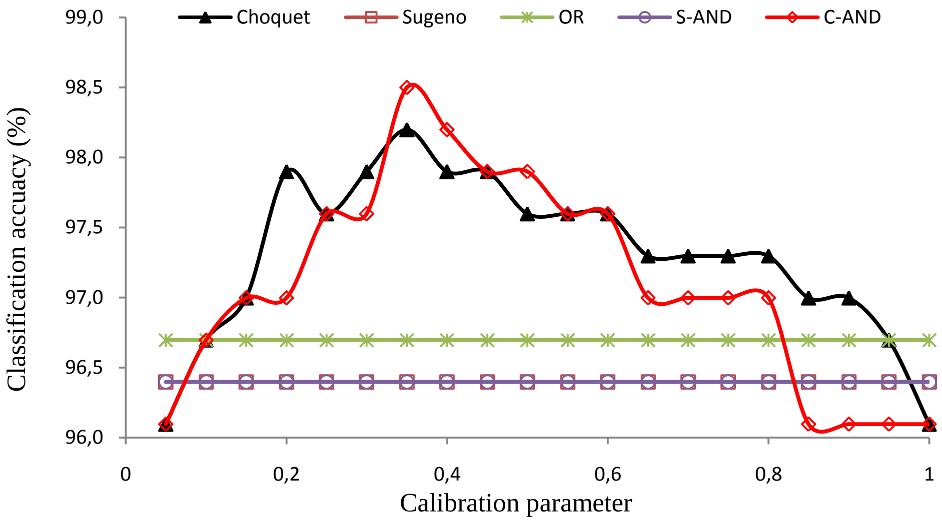
varied in the range [0, 1]. From the presented results, it is worth noting that the Choquet and C-AND are the best and the most suitable FI operators. However, they are highly affected by the calibration parameter δ. Therefore, a careful tuning should be carried in order to achieve the best performance. For instance, for the NN classifier, C-AND operator achieves a classification accuracy from 96.09% to 98.49% when varying the calibration parameter in the range [0.1, 1], which justifies its introduction to the proposed density measure function. Table 4.2 reports the optimal parameters (δ, α, β) selected during the design step to achieve the best classification accuracy.



(a)



(b)



(c)

Figure 4.3: Effect of calibration parameter δ on the different operators: (a) PCA, (b) K-Means and (c) NN classifiers.

Table 4.2: Optimal parameters (δ,α,β) selected for each operator and OCC during the design step.

	Operator		
Classifier	S-AND (δ, α)	C-AND $(\delta, \alpha,)$	OR (δ, α, β)
PCA	(0.05, 0.3)	(0.2, 0.3)	(0.05, 0.2, 0.6)
K-Means	(0.05, 0.5)	(0.5, 0.5)	(0.75, 0.3, 0.5)
NN	(0.05, 0.8)	(0.35, 0.8)	(0.05, 0.6, 0.3)

4.3.4. Combination results

Results for the individual classifiers and the hybrid OCC ensembles with different combination rules according to the selected OCC are reported in tables 4.3. When comparing the individual classifiers, we can note that PCA is the most suitable for this application. Moreover, for a fairly comparison and analysis of the achieved combination schemes, various FI operators are evaluated against fixed combiners including Average (Avg), Product (Prod), Minimum (Min) and Maximum (Max) (Cyganek and Krawczyk, 2015; Duin, 2002). From the obtained results, we can also notice that combining information sources allows improving considerably the classification accuracy than the single source for all classifiers. For instance, when using the hybrid PCA for the test set A, the classification accuracy is improved by more than 7 % against the best single information source, which confirm the effectiveness of using multiple systems than using a single one. Secondly, for the combination strategy, we can clearly observe that the average and product are the best combiners form the fixed group and on the other side, we find that Choquet is the best form the FI group. Besides, when comparing the fixed group against the FI one, we can note that the Choquet and its extension C-AND combiners offer an improved recognition rate than the best fixed aggregators. Therefore, FI combiners are more suitable than fixed ones for hybrid OCC ensemble, Choquet integral seems the most suitable for achieving the hybrid OCC ensemble, since it yields better results with C-AND without any parameter.

Table 4.3: Classification accuracy (%) of individual classifiers and hybrid one-class ensembles with different combination rules for 300 classes.

	Sources			Fixed rules				Fuzzy integral Operators				
OCC ensemble	ES	EM	EA	Avg	Prod	Max	Min	Sugeno	C-AND	OR	Choquet	S-AND
PCA	71.14	73.32	76.26	82.02	82.02	71.11	77.27	79.91	81.86	81.84	**83.73**	78.18
K-Means	60.43	62.65	67.35	76.43	76.32	62.61	69.07	70.51	75.12	71.43	**77.64**	71.88
NN	60.22	62.69	68.15	76.26	76.21	62.64	69.75	72.45	75.67	73.23	**77.55**	73.22

In order to have extended view about the performance of the proposed system, results are provided in table 4.4 with different types for each sort of used OCC to perform the open hybrid OCC ensembles. From the obtained results, we can obviously notice the effect of the selected top for the word recognition. Furthermore, the OC-PCA based open hybrid OCC ensemble appears to perform the best word recognition accuracy, and provides 83.73% and 96.17% for Top-1 and Top-5, respectively.

Table 4.4: Classification accuracy (%) of individual classifiers and hybrid one-class ensembles with different combination rules for 300 classes according to Top-N.

		Sources			Fuzzy integral Operator
	OCC ensemble	ES	EM	EA	Choquet
Top-1	PCA	71.14	73.32	76.26	**83.73**
	K-Means	60.43	62.65	67.35	77.64
	NN	60.22	62.69	68.15	77.55
Top-2	PCA	79.87	80.16	82.88	**87.66**
	K-Means	68.67	68.96	71.22	78.00
	NN	68.15	69.12	71.08	77.62
Top-3	PCA	87..98	90.19	91.97	**92.60**
	K-Means	71.75	72.47	74.32	80.29
	NN	71.22	72.32	74.85	80.16
Top-4	PCA	90.35	91.60	93.25	**95.70**
	K-Means	74.15	74.90	76.74	82.84
	NN	73.85	75.00	76.57	82.78
Top-5	PCA	91.95	92.66	94.88	**96.17**
	K-Means	75.57	76.59	78.38	83.96
	NN	75.48	76.80	78.67	83.73

In order to show the effective use of the Choquet operator against the remaining combination rules, we use the McNemar's test (Dietterich, 1998), which allows comparing statistically two systems. More precisely, a contingency table is constructed in order to calculate the *p*-value.

The McNemar's test has the ability to provide whether one system is significantly better than another according to the *p*-value. A small *p*-value indicates a significant difference of the classification accuracy between two systems to be compared. In contrast, when the *p*-value exceeds 0.05 then the null hypothesis is considered. In this case, both systems perform closely and the difference is too small to decide the superiority of one system than the other.

Tables 4.5 reports the *p*-value of the Choquet operator against the remaining combination rules for PCA, K-Means and NN classifiers according to Top-1. The obtained results show that the Choquet operator is significantly different from the other combination rules for the different type of classifiers except for C-AND operator, where small difference can be observed in some cases for which *p*-values exceed 0.05.

McNemar's test proves that combination rules based on Choquet and C-AND are more robust than the fixed rules whatever the used OCC and datasets. Indeed, the obtained *p*-values are very small which shows the high difference between the different aggregators.

The obtained results show the successful of Choquet operator for improving the results of the hybrid OCC ensemble, which is due to its ability to capture interactions among the OCCs and attribute the right importance for each information source. According to the definition of Choquet FI, the appropriate weight values are deduced from the fuzzy measure dynamically for each test pattern which represents its successful key. Moreover, the proposed method for dynamic density measure, which is associated with fuzzy measure, offers a better adaptation allowing to give more importance to the relevant information source relatively to the others for each test sample.

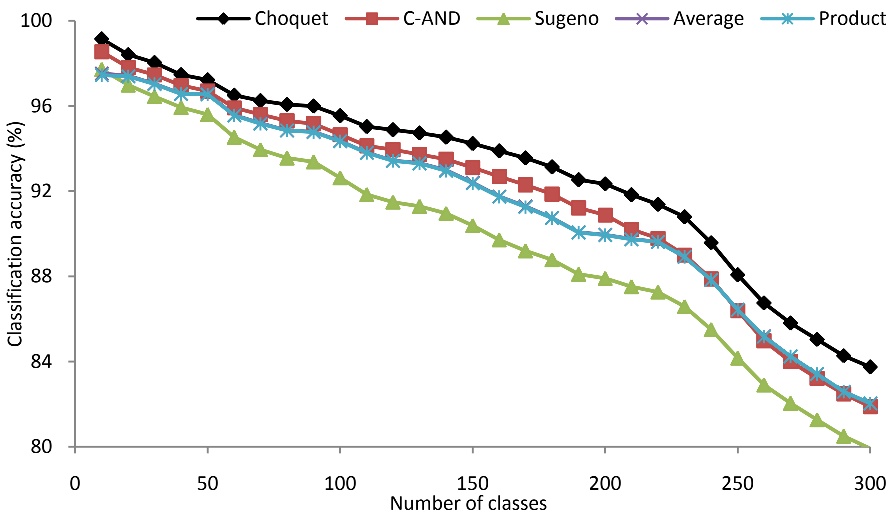
Table 4.5: The p-values of McNemar's test for Choquet FI versus the other combination rules for hybrid OCC ensembles according to Top-1.

	Fixed Rules				Fuzzy Integral Operators			
OCC ensemble	Avg	Prod	Max	Min	C-AND	OR	Sugeno	S-AND
PCA	$8.28\ 10^{-13}$	$2.28\ 10^{-13}$	$1.17\ 10^{-8}$	$< 10^{-16}$	$< 10^{-16}$	$< 10^{-16}$	$< 10^{-16}$	$< 10^{-16}$
NN	$1.50\ 10^{-7}$	$2.16\ 10^{-8}$	$< 10^{-16}$	$< 10^{-16}$	$3.50\ 10^{-5}$	$9.74\ 10^{-14}$	$< 10^{-16}$	$< 10^{-16}$
K-Means	1.77×10^{-6}	2.01×10^{-7}	$< 10^{-16}$	$< 10^{-16}$	$4.28\ 10^{-5}$	$< 10^{-16}$	$< 10^{-16}$	$< 10^{-16}$

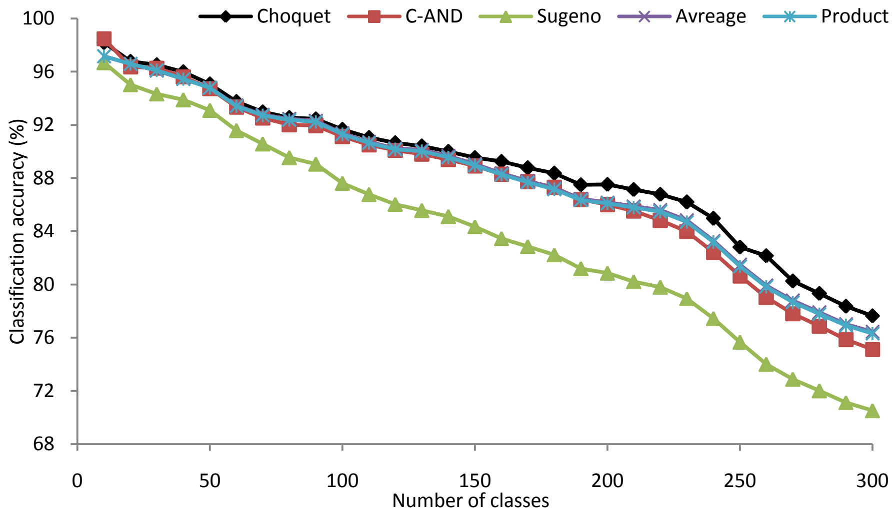
4.3.5. Stability study of combination rules

The proposed hybrid OCC ensembles using FI operators are evaluated to show the behavior of the parameter-independent open classification when new classes are progressively added to the system. Also, we use the stability criterion, which defines the ability of an OCC or hybrid

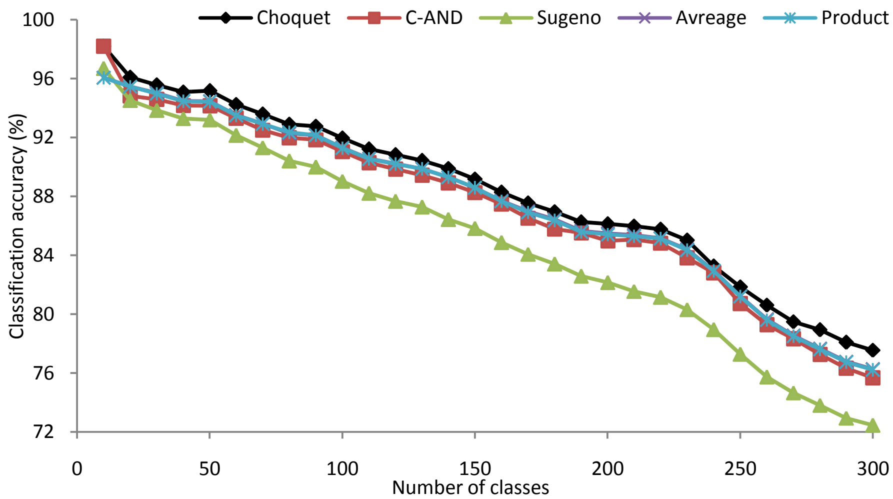
OCC ensemble to maintain the same classification accuracy when the number of classes is progressively added to the system. Hence, the classification accuracy is computed for the hybrid OCC ensemble by adding progressively new classes each time from 10 to 300 using parameters found in the design step. Figure 4.4 depicts the classification accuracy versus the number of classes achieved by the choquet, C-AND and Sugeno operators against best fixed rules average and product to perform the hybrid OCC ensembles. Roughly speaking, we clearly can notice that the most combiners achieve similar classification accuracy for few classes. However, when adding progressively new classes, the used combination methods behave differently for more complicated problem. When comparing the set of aggregators, the Choquet is the most stable and widely better than the fixed combiners since they keep the same classification accuracy while adding new classes. Indeed, the gap between the best and the worst aggregators increases progressively when extending the hybrid OCC ensemble for more complex problem via adding new classes. Consequently, the Choquet operator provides the best performance and are the least affected when adding new classes. Conversely, surprising results are achieved by Sugeno FI, which is highly affected by the number of classes and shows its inappropriate use for combining hybrid OCC ensembles.



(a)



(b)



(c)

Figure 4.4: Stability study of combination rules for hybrid one-class ensembles performed on three OCCs: (a) PCA (b) Kmeans (c) 1NN.

4.3.6. Discussions

Handwritten Arabic word recognition finds its interesting use in different real world applications. HMM, SVM and neural networks are the most used classifiers for Arabic word recognition, however, these classifiers provide an Arabic word recognition system with closed lexicon. Nowadays, extended multi-class implementation to new classes is strongly required, for instance, in Arabic word recognition and handwritten writer identification. Indeed, the use of OCCs for solving the multi-class classification problem has been discussed by many researchers due to its offered properties. This type of classifier attempts to model each class separately from the others that allows designing an open multi-class system. This property is highly desirable for the actual systems since it is possible to add new classes without re-training the classification system again on the whole classes. However, using individual OCC does not allow designing a robust classification system. Therefore, the design of hybrid OCC ensemble is required in order to achieve the best possible classification performance and robustness as well as keeping the properties of OCC. Thus, this chapter discusses the appropriate use of the FI for best designing of the open-lexicon handwritten Arabic word recognition system. Indeed, various works reported that FI provides excellent results for classifier aggregating. However, when adapting FI for the hybrid OCC ensemble, a problem is faced which is the generation of the density measure. Indeed, it is usually estimated using the training datasets. However, this estimation is considered not representative and leads to require new validation dataset to get a better evaluation of each information source. Furthermore, all testing samples are represented by the same density values, which make this approach less efficient since each test sample has its suitable density measure. Therefore, more contribution should be provided in respect to the more appropriate information source and each test sample. To overcome this drawback, a dynamic density measure is proposed having the ability to attribute the appropriate values to each information source as well as its adaptation for each test pattern.

In order to design a parameter independent system, a new framework is proposed that relies on using a separated datasets for tuning parameters and the selection of optimal parameter values. This is an efficient property since, once the parameters are found, they are considered to be the same for all existing classes and also for the new ones.

In this step, we have seen the effect of all parameters including descriptor, classifier, density measure and OWA operators. From the set of parameters, we notably mention the impact of the calibration parameter used in the density measure to enhance the combination performance. It can be also noticed that the calibration parameter is more suitable for the Choquet and C-AND operators. From the obtained results we clearly notice the achieved improvement by the use of the FI combination scheme based on Choquet operators against the other operators and fixed rules. This superiority can be verified in terms of the classification accuracy and *p*-values. More precisely, the Choquet FI shows its stability to maintain the classification accuracy roughly stable when adding new words to the lexicon against the other combination rules. However, it is difficult to compare the proposed system against the state-of-the-art systems, due to the fact that we offer an additional advantage. Finally we conclude that, in addition to the open-lexicon propriety in the presented system, the obtained results are promising.

4.4. Conclusion

By this chapter we aimed to study the usefulness of the hybrid OCC ensembles to perform open-lexicon handwritten Arabic word recognition. Moreover, due to the fact that fixed rules are the standard used combiners for the hybrid OCC ensemble, the proposed work attempted to study the potential of FI operators *via* proposing a combination scheme for combining ensemble of OCCs designed by different feature generation methods.

Experimental results are conducted on different types of OCC and Arabic handwritten word datasets having high number of classes. Obtained results prove the superiority of FI against fixed combiners those represent our base of comparison, whatever the selected type of OCC. Furthermore, the Choquet operator seems to be the most suitable and powerful among FI aggregators. Thus, this study suggests keeping fuzzy integral operators as a way for achieving robust hybrid OCC ensembles for performing open-lexicon handwritten Arabic word recognition.

According to the obtained results, in the continuation of this chapter we will focus on performing the best hybrid OCC ensemble based on PCA classifier and Choquet FI, for writer identification application based on handwritten signature.

Chapter 5

Hybrid One-class PCA Ensembles and Curvelet Transform for an Efficient Open Offline Handwritten Signature Identification System

Abstract

In this chapter, an Open Handwritten Signature Identification System (OHSIS) is proposed by using conjointly the Curvelet Transform (CT) and the One-Class classifier based on Principal Component Analysis (OC-PCA). CT is explored for feature generation due to its efficient characterization of curves contained into the local orientations within the signature image. While, OC-PCA is used for its effectiveness to absorb the high feature size generated by the CT and allows achieving at the same time an open system. Then, in order to improve the robustness of the OHSIS when few reference signatures are available, a combination based on Choquet fuzzy integral is explored to combine multiple individual OHSISs. Furthermore, a designing protocol with limited number of writers and reference signatures is proposed to perform a parameter-independent OHSIS. Experimental results conducted on standard CEDAR and GPDS handwritten signature datasets report 97.99% and 94.96% correct identification rate, respectively, which highlights the effectiveness of the proposed OHSIS since it can comfortably outperform the state-of-the-art when using few reference signatures.

5.1. Introduction

Biometric recognition is nowadays considered as an active research field due to its effective use in law enforcement, forensic sciences as well as its increasing requirement in wide variety of civilian applications for enhanced security and privacy (Jain et al., 2004). Biometric recognition is defined as an automatic identification and verification of persons. Therefore, various biometric modalities have been explored in the last couple of decades (Vielhauer and Dittmann, 2006; Boyer et al., 2007), which are based on physiological and behavioral characteristics. Physiological characteristics are related to anatomical properties of a person including for instance face, fingerprint, iris and hand geometry. In contrast, behavioral characteristics illustrate skilled action performed by a person such as the voice, handwritten signature and gait.

Handwritten signature is considered as the oldest and the most widely accepted biometric tool for person identification and verification. Signature identification aims to find the writer's identity within database among the set of writers. In this case, the questioned signature (Pavlidis et al., 1998) is compared to all writer models contained in the database. In contrast, the purpose of the signature verification system is to decide whether a questioned signature truly belongs to a person or not. In this case, this signature is only compared with the writers' model. Hence, the state-of-art reveals great efforts for proposing various algorithms in order to achieve efficient signature verification (Srihari et al., 2004; Bertolini et al., 2010; Kumar et al., 2012; Elrajubi and El-Feghi, 2015). However, signature identification has received less attention, despite its interesting in practical application areas. For instance, companies validate the identity of each individual for accessing to certain security-sensitive facilities (Martinez and Sanchez, 2006). Another interesting use of signature identification is in law-enforcement applications, where the identification of perpetrators is a fundamental requirement of the solution, in the analysis of some historical documents and automatic bank check processing (Ismail and Gad, 2000).

Handwritten signature identification can be performed using two acquisition modes: the online mode and the offline mode. The design of the handwritten signature identification systems based on the offline mode is more difficult comparatively to the online mode since many desirable characteristics such as the velocity and the pressure are not available during the acquisition. The identification system depends only on the feature selected from the signature shape (Pirlo, 1994).

So far, the identification systems provide satisfactory results when the number of reference signatures per writer is large (Pavlidis et al., 1998; Bertolini et al., 2010; Martinez and

Sanchez, 2006; Ismail and Gad, 2000 Han and Sethi, 1996; Riba et al., 2000; Kalera et al., 2004). However, when the number of reference signatures is small, the writer identification still awaits proper attention to get more improvements in terms of accuracy and robustness. Indeed, in a real environment, each writer will be willing to provide a limited number of original signatures. In this case, the system should be carefully designed using only few reference signatures to get a relevant identification. Moreover, the identification system should be flexible as much as possible to be extended for new writers without neither retraining nor recalculating the optimal parameters. Thus, the present work aims to develop an efficient Open Handwritten Identification System (OHSIS) via contributing various stages, which are feature generation, classification and combination of multiple OHSIS. For feature generation, the Curvelet Transform (CT) (Candès and Donoho, 1999) is investigated for its effective use in variety of applications in pattern recognition and image compression (Feng et al., 2011; Li et al., 2009). Besides, CT has been successfully used for offline handwritten signature verification. In addition, to allow an open identification which deals with prospective new writers, the classification is generally performed through the use of distance classifier or matching process (Han and Sethi, 1996; Riba et al., 2000). Indeed, robust classifiers such as Bi-SVM and MLP achieve closed systems, that are able to identify only the writers contributing in the enrollement stage (Martinez and Sanchez, 2006). In other words, when a new writer is added to the system, all the classifier should be retrained again on the whole data. Consequently, a such system is difficult to deploy for instance in the bank.

Recently, One Class Classifier (OCC) has attracted much attention to many researchers for solving the multi-class classification problem (Cyganek, 2000; Goh et al., 2005; Ban and Abe, 2006; Rabaoui et al., 2008 ; Yeh et al., 2009; Boehm et al., 2011), owing to its offered advantages against the usual implementations based on binary or multi-class classifiers. Indeed, when a new writer is added to the HSIS, the used OCCs do not require retraining for a second time. Besides, the OCC offers less computational cost in terms of training time and memory space against some multi-class implementations (Hsu et al., 2002) such as OAO and OAA based on Bi-SVM classifiers (Yeh et al., 2009). Presently, this work tries to investigate the applicability of the OCC for signature identification. More precisely, this chapter proposes to investigate the use of OC-PCA from various available types of OCCs (Tax, 2001) for its ability to absorb the high size of the feature vector generated by the CT and allows generating a well representative model for each writer (Hadjadji et al., 2017b).

Furthermore, in a real environment, a user will be willing to provide a few signatures, which can highly affect the performance of the OHSIS. Hence, in order to achieve more efficient OHSIS, a combination of multiple OHSIS is proposed by means of a new density estimation associated to the well-known Choquet FI. Usually, when extending the OHSIS for new writers, parameters are required to be found for the added ones. To overcome this drawback, this work proposes an alternative protocol to design a parameter-independent OHSIS.

The remaining of this chapter is organized as follows. Section 2 presents an overview of the two tools used for designing the OHSIS, which are based on CT and the OC-PCA classifier. Section 3 describes in details the proposed individual and combined OHSIS. In order to evaluate the effective use of the proposed approach, experimental results conducted on the well known datasets are presented in section 4. Finally, the conclusion and future work are provided in the last section.

5.2. Related works

Generally, the offline handwritten identification system is composed of two main modules: feature generation and classification. Thus, many attempts have been done for improving the offline handwritten identification system by addressing the feature extraction and matching stages. For the feature extraction, one of the oldest identification system was proposed by (Han and Sethi, 1996), in which a set of geometrical and topological features are considered to map a signature image into two strings of finite symbols. The geometric features include horizontal bars, vertical bars and loops extracted from the skeleton image of the signature using the 4-connected component labeling algorithm. The horizontal and vertical bars are extracted from the binary image of the signature using the morphological hit-or-miss operation. The set of topologic features consists of end points, branch points, crossing points, convex points and concave points. Furthermore, (Pavlidis et al., 1998) proposed the use of a revolving active deformable model for capturing the unique characteristics of the overall signature structure. In addition, a polygonal approximation algorithm is applied to smooth excessive information gathered at certain signature parts (e.g., like small fluctuations of almost straight lines, which might be of some value for verification purposes, but are rather harmful for identification purposes). In another work, (Ismail and Gad 2000) followed this successful path to improve the handwritten signature identification by introducing a suitable combination of distinctive and effective global features (e.g. width and the baseline) and local features (e.g. critical points and gradients). In another attempt, (Riba et al., 2000) proposed a feature generation method based on different types of moments, and then a canonical variable analysis is carried out in order to reduce the number of features. In related work, (Kalera et al., 2004) proposed a novel approach for offline handwritten signature verification and identification based on a quasi-multiresolution technique using a combination of gradient, structural and concavity features. Recently, (Martinez et al., 2007) addressed the problem of offline handwritten signature identification *via* using a feature vector constructed from global geometric and moment-based characteristics. A mechanism to capture the intrapersonal variability of each user using just one original signature has been presented. More recently, (Sulong et al., 2014) proposed the use of the adaptive window positioning technique as an efficient feature extraction for offline handwritten signature identification. The proposed technique mainly, employ the division of signature images into 13x13 windows, where this size should be large enough to contain ample information about the style of the author and small enough to ensure a good identification performance. In addition, this technique creates some new cluster patterns for each window when classified into groups of similar attributes. Experimental results reveal that adaptive window positioning technique proved to be the efficient and reliable method for accurate signature feature extraction for the identification of offline handwritten signatures.

For the classification stage, various matching algorithms have been proposed. For instance, (Han and Sethi, 1996), used a 2D string for matching two signatures in order to take into account the 2-dimensional spatial relationships of the different features contained in a signature. The Longest Common Subsequence (LCS) matching criterion is used as a measurement of similarity between the questioned signature and a reference one, for the retrieval and identification. The reference signature yielding the longest LCS is considered as the most similar to the questioned one. Furthermore, (Pavlidis et al., 1998) proposed a novel Synchronized String Matcher (SSM) algorithm in order to match accurately the questioned and reference signatures, which is inspired from the error recovery techniques in compiler design. Basically, the SSM algorithm tries to resynchronize the matching process between the reference signature string and the questioned signature string. In another work, (Ismail and Gad, 2000) developed a multistage classifier to perform the recognition step in three stages. First, a pre-classification stage is applied to group similar slant signatures. Then, an identification scheme is proposed to discriminate between individuals within a group. In the second stage, distances between the global feature vector of the input questioned sample and the mean of each class in the group are computed and compared sequentially in order to select the best three candidates. Finally, in the third stage, local center points are used to whether select the best candidate or to reject the questioned signature based on a threshold value of each candidate class. In a similar work, (Riba et al., 2000) addressed the classification stage through a comparison of several statistical methods. The obtained results show that the linear discriminant analysis performs better in terms of accuracy and computational cost than other classifiers, such as principal components regression, partial least square, quadratic discriminant analysis, k-Nearest Neighbor (k-NN), fuzzy logic and two neural networks methods including *backpropagation* and the radial basis model. Thereafter, (Kalera et al., 2004) demonstrated the usefulness of statistical measures like the Bayes and k-NN classifier. Recently, (Martinez et al., 2007) addressed the problem of classification stage by comparing the performance of the well-known Binary Support Vector Machines (Bi-SVM) and Multi-Layer Perceptrons (MLP). The experimental results evince the superiority of the Bi-SVM classifier.

5.3. Design of OHSIS

In order to design an open system, M writers are considered as available for finding the optimal parameters of the OHSIS.

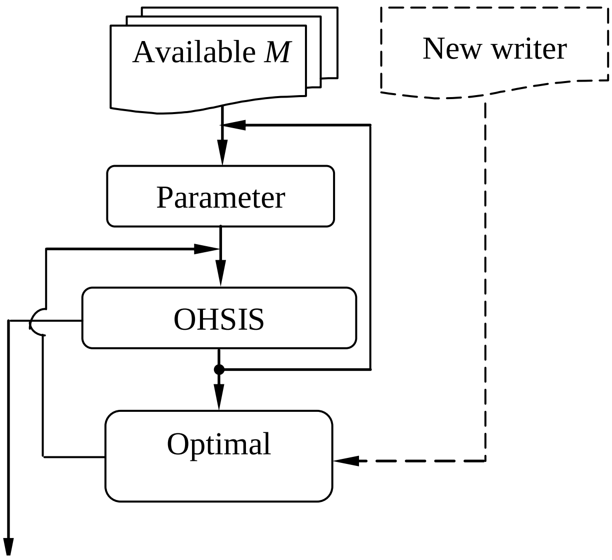


Figure 5.1: Design of the overall OHSIS and its extension for new writers.

As depicted in figure 5.1, from the available writers, parameters of the OHSIS are adjusted successively till finding the optimal values using a criterion based on the maximal prediction.

When a new writer is added to the OHSIS, the optimal parameters are assigned for him. Thus, when using the OCC, the scheme has the advantage to add a new writer directly into the OHSIS having its own parameters without retraining the overall system as it is required when using other classifiers such as Bi-SVM or MLP (Martinez et al., 2007). In the following, the individual as well as the combined OHSIS are described more precisely for writer identification.

5.3.1. Individual OHSIS

The open identification system as shown in figure 5.2 is composed of three main stages, which are signature preprocessing, feature generation and classification. In the following, each stage of the proposed OHSIS is described for writer identification.

5.3.1.1. Preprocessing

Binarization is the main preprocessing of the acquired signature. In this work, the local iterative threshold is used (Ridler and Calvard, 1978), which is performed recursively for each sample area found in a square centered on the pixel at 9x9. The mean and the standard deviation are calculated for each pixel in the sample area. Therefore, these values are used for deciding if the current pixel takes 0 or 1.

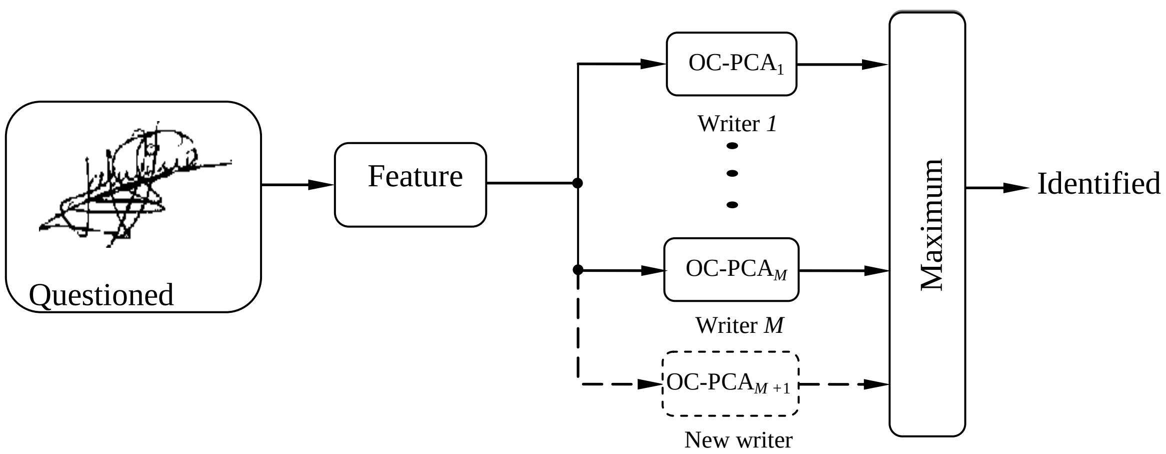


Figure 5.2: Individual OHSIS scheme.

5.3.1.2. Curvelet transform-based feature generation

CT has been developed specifically for representing curves and other singularities along curves contained into an image using fewer coefficients for a given accuracy of reconstruction. In the proposed system, CT is used for generating the feature vector from the signature image *via* the wrapping technique at different scales and orientations.

The main advantage of the CT is its ability to capture curves into the handwritten signature. In order to capture more efficiently the local information, CT is performed on a grid over the signature image. Thus, two different ways are proposed for feature generation by partitioning the signature image using equi-space and equi-mass grids as it is depicted in figure 5.3. A uniform or equi-space grid (Favata and Srikantan, 1996) creates rectangular regions for sampling, where each one has the same size and shape. Conversely, an equi-mass grid creates different rectangular regions having the same number of black pixels, also known as the mass, of the word image grid (Favata and Srikantan, 1996). Thereby, each region is found by partitioning horizontally and vertically the signature image using its mass histogram. Hence, the total mass between all adjacent regions are as close to be equal as possible.

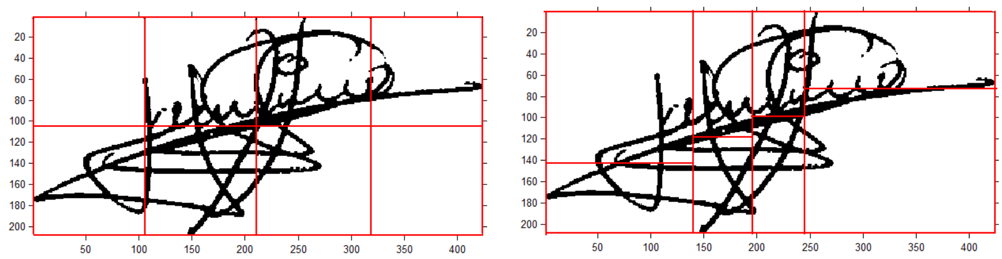
The proposed feature generation of each signature is then performed through the following steps:

1. Divide the signature image into several cells *via* equi-space or equi-mass grid.
2. Normalize each sub-image size to $N \times N$ by adding the ones around it, which is the signature background as it is required for the CT calculation.
3. Perform the wrapping CT for each sub-image leading to generate curvelet coefficients at multiple scales and orientations.
4. Calculate the energy E at each scale j and orientation k as:

$$E(j,k) = \sum_{t_1} \sum_{t_2} \left| C_{j,k}(t_1, t_2) \right| \tag{5.1}$$

   $C_{j,k}$ defines the curvelet coefficients computed at the scale j and the orientation k.

5. Concatenate all energies computed from the different sub-images to generate the feature vector.



(a) (b)

Figure 5.3: Example of equi-spaced (a) and equi-mass adaptive grid (b) grids with 2×4

5.3.1.3. OC-PCA for OHSIS

Classification is the last step for writer identification, which consists to assign a feature vector of a questioned signature to one of the writers. Let $C = \{c_1, \ldots, c_M\}$ be a set of M writers each one has its corresponding reconstruction error f^{PCA} associated to each $\mathrm{OC{-}PCA}_m$ $(m = 1, .., M)$. After computing the writer model in the training step, a questioned signature is assigned to its corresponding writer that generates the maximum prediction. The class label $y(x)$ of the questioned signature x is defined as follows:

$$y(x) = \underset{m=1}{\overset{M}{argmax}}(f^{PCA}(x)) \tag{5.2}$$

Thus, the identification system is composed of a number of OC-PCA, which is equal to the number of writers defined in the database. When a questioned signature is presented to OHSIS, it is assigned to the writer with the best matching between the OC-PCA model with the signature feature vector.

Note that, it is not needed to define the threshold values for multi-class implementation and consequently all thresholds are set to zero.

5.3.2. Choquet integral-based multiple OHSIS

The design of an individual OHSIS is effective when a large number of reference signatures is available. However, in practical applications as in the bank, one OHSIS offers medium performance when a limited number of signatures is available since, in a real environment, a user will be willing to provide only 3 to 5 signatures. Therefore, a simple and successful way to improve the identification scores is to combine multiple OHSIS allowing to take advantage of each individual OHSIS. Thus, a great effort has been done for proposing various combination methods and schemes including methods based on fuzzy operators. In this respect, FI and the associated fuzzy measures initially introduced by Sugeno are reported to give excellent results for classifier aggregating. Its main advantage is to measure the strength not separately for each classifier alone but for all members. The ability of the fuzzy integral to enhance the results produced by multiple information sources has been highlighted in various application areas of pattern recognition (Kwak and Pedrycz, 2005; Chiang and Gaber, 1997; Pham and Wagner, 2000; Chiang, 1999; Cho and Kim, 1995; Cho, 1995). Thus, the contribution of the Choquet Fuzzy Integral (C-FI) is investigated in order to achieve a robust signature identification system. The mathematical foundation and the application of C-FI are detailed in the next sections.

The combined OHSIS as it is depicted in figure 5.4 is composed of M writers and 2 different information sources generated by means of the CT. Consequently, two different classifier models are generated for each writer using the OC-PCA.

Several combination rules are possible to achieve the aggregation, but all these rules need a unique interpretation of the outputs generated by the different classifiers for each questioned signature x. Hence, the normalization of outputs for each classifier is required to perform correctly the combination. Let M be the set of single OCC ensembles and f_m^i the reconstruction error of the OC-PCA$_m^i$ trained on the i^{th} information source of the m^{th} writer. Hence, an exponential function is used for transforming the OC-PCA output f_m^i ranging from the interval $]-\infty, 0]$ into posteriori probability $P_i(c_m/x)$ ranged in the interval]0, 1] expressed as follows:

$$P_i(c_m/x) = exp\left(f_m^i(x)\right) \tag{5.3}$$

The evidence is then expressed as the posteriori probability taking the following form:

$$h_m(z_i) = P_i(c_m/x) \tag{5.4}$$

As already introduced, the successful key of FI strongly depends on the appropriate choice of the density measure. Consequently, if density measures are well defined then the fuzzy measures can be correctly identified, which makes it suitable for the operation of the fuzzy integral.

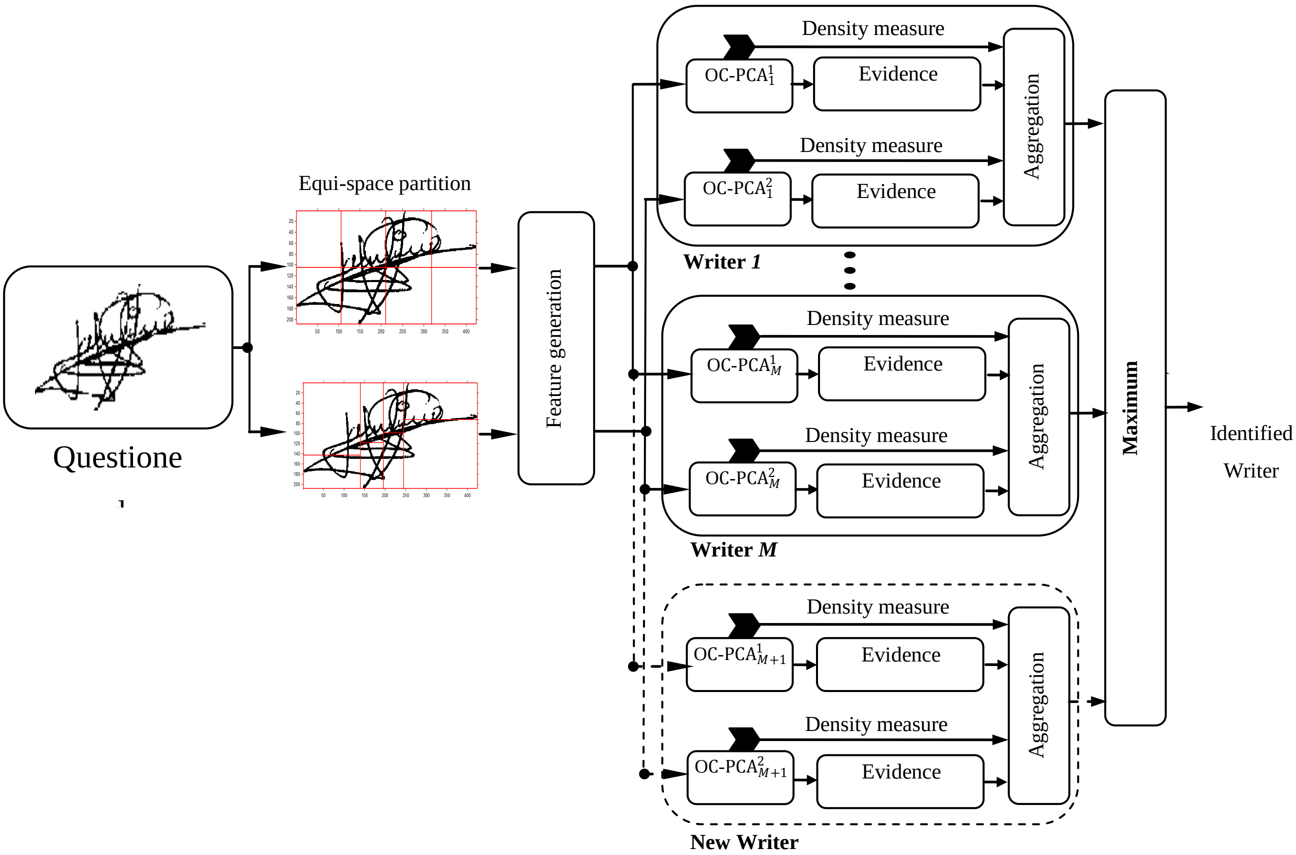


Figure 5.4: Combined OHSIS scheme based on Choquet fuzzy integral.

Thus, in this chapter we will explore the already proposed technique, in which the density measure is considered as the similarity degree or correspondence between the writer model and that of the questioned signature, i.e. the closer the similarity is, the greater the value of the density measure is. Precisely, the density measure is defined dynamically for each questioned signature as:

$$g_m^i(x) = exp\left(-\delta \left\| f_m^i(x) - \bar{f}_m^i \right\|^2\right) \tag{5.6}$$

Such that, $0 \leq g_m^i(x) \leq 1$ and $0 < \delta \leq 1$.

$\bar{f}_m^i$ is obtained by averaging outputs f_m^i using the training signatures. δ is a positive value that has been introduced for calibrating the density till finding the most suitable value for representing the contribution of each classifier and information source. In addition, the δ parameter helps to match the density with normalized outputs allowing a better calibration and therefore more improvement of the accuracy.

After getting the densities, the fuzzy measures, $g_m(A_i)$ for $1 \leq i \leq L$ and $1 \leq m \leq M$, are calculated by applying Eqs. 1.24 and 1.25.

Let Ic_m as the Choquet aggregation produced from the posteriori probabilities of the different information sources and their corresponding fuzzy measures defined for each writer, then the questioned signature x is assigned to the writer class with the maximal value as follows:

$$y(x) = \underset{m=1}{\overset{M}{argmax}} \left(Ic_m(x)\right) \tag{5.7}$$

Thus, the OHSIS is composed of a number of OC-PCA ensembles each one is aggregated by the Choquet FI, which is equal to the number of writers defined in the database. When a questioned signature is presented to OHSIS, it is assigned to the writer best matching the OC-PCA ensemble model with the signature feature vector.

5.4. Experimental results

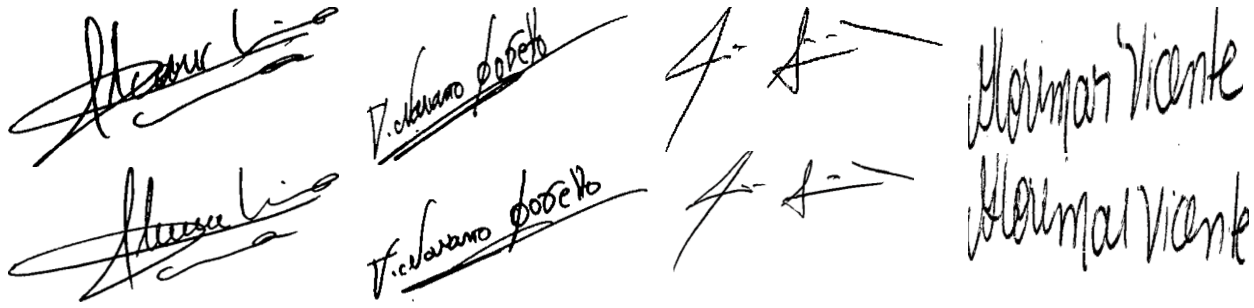
5.4.1. Datasets and evaluation criteria

Two publicly available offline signature datasets are used for evaluating the performance of the proposed OHSIS. The first one is the well known "Grupo de Procesado Digital de Senales" (GPDS) signature dataset (Vargas et al., 2007) containing 300 writers each one has 24 genuine and 30 forgery signatures, respectively. While the second one is the CEDAR signature dataset containing 55 writers (Kalera et al., 2004), each one has 24 genuine and 24 forgery signatures. Figure 5.5 shows two different genuine signatures for two different writers. For evaluating the performance of OHSIS, only genuine signatures are taken into consideration to calculate the Identification Rate (IR) expressed in %, which is defined as the number of instances correctly identified to the total number of instances formally formulated as:

$$\text{IR} = 100 \times \frac{\text{Number of correct identification for all writers}}{\text{Total number of test signatures for all writers}} \ (\%) \tag{5.8}$$

During the evaluation, GPDS dataset is used for designing and evaluating the OHSIS, while CEDAR dataset is used only for testing the OHSIS designed from GPDS dataset.

In this work, since only the handwritten signature identification problem is addressed, therefore there is no need to use the forged signatures. In addition, all questioned signatures belonging to the writers are assumed in the database and there is no need to reject anyone.



(a) **(b)**

Figure 5.5: Handwritten signature samples from GPDS (a) and CEDAR datasets (b)

5.4.2. Experimental setup

The proposed OHSIS is developed through design and identification steps. During the design step, a sub-set of writers is selected from the whole dataset for finding the optimal parameters of the feature vectors and OC-PCA classifiers. During the identification step, the OHSIS involves testing with the remaining writers having the same parameters found during the design step.

Generally, 10-fold cross validation is the most used protocol for evaluating the pattern recognition system (Kohavi, 1995). However, for applications such as handwritten signature identification, a limited number of signatures is available for training purposes. Indeed, this characteristic is very important and should be taken into consideration for designing a practical OHSIS, since, in a real environment, a user will be willing to provide a limited number of original signatures (Martinez and Sanchez, 2006). Hence, during the design step, some signature samples per writer are randomly selected in order to train the OC-PCA classifier. Usually, three to five reference signature specimens are required for enrolling the writer to the HSIS. For a strict requirement, only two reference signatures are used for designing the OHSIS.

Hence, for designing parameter-independent OHSIS, the GPDS dataset is divided into two sub-sets, one is used for designing the OHSIS and the remaining is used for evaluating its robustness. In this context, only 10 writers are assumed available initially for training OC-PCA classifiers, each one is trained using only 2 signatures. The 22 remaining signatures per writer are used for finding the best parameters of the grid size for generating features and the number of eigenvectors for the OC-PCA. Thus, signatures of 10 randomly selected writers are used for finding the suitable parameters to best design the OHSIS. When a new writer is presented to the system, the same parameters are used. In this case, simply provide some reference signatures for its design without recalculating the parameters.

As reported previously, two different ways are proposed for feature generation, which are equi-space and equi-mass. Both methods are highly affected while adjusting the grid size by changing the numbers of rows (Nr) and columns (Nc), respectively. Thus, both numbers are experimentally fixed. Specifically, 2 signatures per writer are used for training and 22 remaining signatures are used for testing in order to find the best grid size. Meanwhile, the OC-PCA classifier parameter should also be carefully tuned. This classifier needs only one parameter that must be specified, which is the number of eigenvectors.

IR expressed in % for the equi-space and equi-mass partitions with various grid sizes obtained with the best classifier parameters are reported in the tables 5.1 and 5.2, respectively.

The obtained results clearly reflect the effect of partitioning the image signature into different partitions for a better handwritten signature characterization. Consequently, adjusting the grid size is important since it influences the identification accuracy. For instance, the IR varies from 84.09% to 93.18% and 87.27% to 96.36 when varying the equi-space and equi-mass grid sizes, respectively. Besides, these results justify the use of the grid size parameter and its careful tune for achieving an efficient OHSIS. As can be seen, the optimal grid size parameters are 2×7 for both partition methods. These selected parameters lead to generate a feature vector having 672 components for both feature generation methods. In addition, the single parameter of OC-PCA is adjusted to 15 eigenvectors for achieving the best IR.

Table 5.1: IR (%) of OHSIS design using 10 writers with various equi-spaced grid sizes. Nr\Nc define the numbers of row and column for the used grids, respectively.

$Nr\backslash Nc$	2	3	4	5	6	7	8
1	84.09	85.00	85.45	86.36	87.73	87.73	89.09
2	90.00	90.91	91.81	91.81	92.27	**93.18**	92.72
3	90.45	91.36	91.36	92.72	93.18	92.72	92.72

Table 5.2: IR (%) of OHSIS design using 10 writers with various equi-mass grid sizes. Nr\Nc define the numbers of row and column for the used grids, respectively.

$Nr\backslash Nc$	2	3	4	5	6	7	8
1	87.27	88.63	90.90	91.36	92.27	93.18	93.18
2	93.18	94.09	91.81	94.54	95.45	**96.36**	94.09
3	94.09	94.54	94.54	95.90	96.36	96.36	95.45

5.4.3. OC-PCA versus Bi-SVM for handwritten signature identification

This test aims to evaluate the performance of OC-PCA against Bi-SVM classifier for achieving the identification system. For a fair comparison, the design step for Bi-SVM is performed to find its parameters (C,γ) using the same grid size as used when performing the OC-PCA. Therefore, both classifiers are examined under the same experimental protocol using 2 reference signatures per writer. In addition, the OAA implementation is used in order to achieve multi-class implementation based on Bi-SVM classifier (Martinez and Sanchez, 2006).

The evaluation of the individual system is performed by adding progressively writers using the same parameters as found during the design step. Hence, the IR is calculated at each time when adding progressively new writers from 10 to 300. Figures 5.6 and 5.7 report the IR for the OC-PCA versus Bi-SVM using Equi-Spaced (ES) and Equi-Mass (EM) partitions, respectively.

It can be clearly observed that when adding progressively new writers, there is natural gradual decrease in the IR notably after 100 writers. This behavior is due to the fact that the identification problem becomes more complex when the number of writers is large and it becomes more difficult to separate between writers having similar shapes of signatures. Moreover, the OC-PCA achieving the best IR is also the most stable and exhibits the least variation in IR when increasing the number of writers. Furthermore, when comparing the OC-PCA versus Bi-SVM results, the OC-PCA achieves better IR against the Bi-SVM classifier for both feature generation methods.

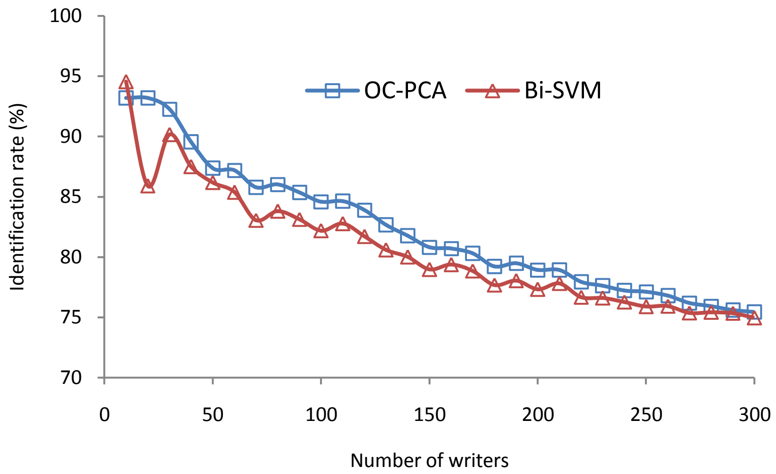


Figure 5.6: Writer IR (%) of individual ES system using OC-PCA and Bi-SVM trained with 2 reference signatures per writer.

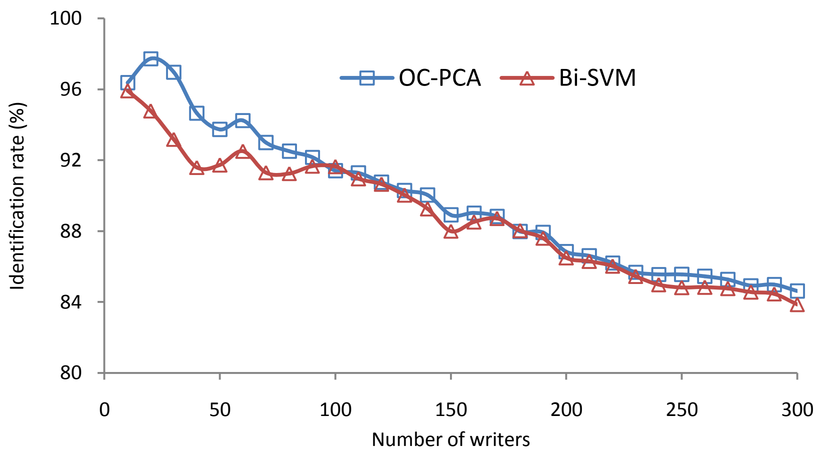


Figure 5.7: Writer IR (%) of individual EM system using OC-PCA and Bi-SVM trained with 2 reference signatures per writer.

Besides, another evaluation criterion that justifies the use of the OC-PCA rather than Bi-SVM is the training time. Thus, figure 5.8 depicts the time reduced ratio performed by the OC-PCA against Bi-SVM classifier versus the number of writers included progressively to the identification system. It is easy to observe the high computational cost reduced by OC-PCA against the Bi-SVM, especially for a large number of writers. Consequently, more the number of writers is increased, more the time reduced ratio is important.

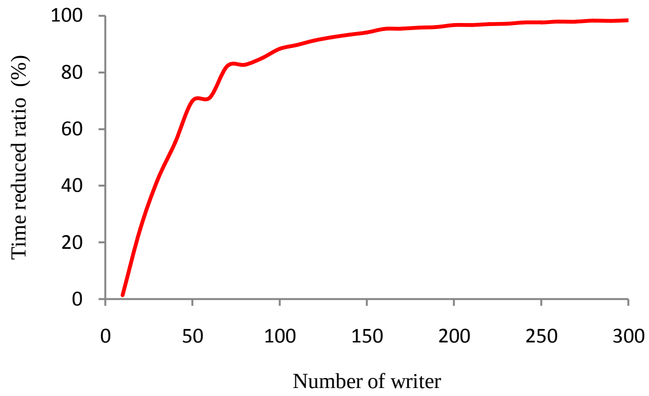


Figure 5.8: Training time reduced ratio by the OC-PCA against Bi-SVM classifier.

Finally, the proposed OC-PCA is more efficient than the Bi-SVM in terms of the IR and training time. Moreover, the use of OC-PCA offers faster and easier extension than the Bi-SVM classifier.

5.4.4. Evaluation of OHSIS combination

In order to improve the accuracy and the robustness of the identification system, individual OHSIS based on ES and EM based descriptors are combined using the Choquet fuzzy integral. However, this combination performs well if individual systems are complementary. Thus, a diversity measure of the individual systems is useful for evaluating their complementarities and therefore to check their abilities to achieve an improved system. The following sections highlight a way for measuring the diversity of two systems as well as the design and evaluation of their combination.

5.4.4.1. Diversity measure

To achieve an useful combination of individual systems, the complementary performance is exploited in order to enhance the IR. Otherwise, the overall decision will not be better than the individual ones. Thus, the difference between two complementary systems or methods is assessed through a measure called diversity. Most used pairwise diversity methods are the Q-statistic, correlation, disagreement and double fault. More details of these methods can be found in (Shipp and Kuncheva, 2002). Two systems can be considered diverse when Q-Statistic and correlation are reduced while disagreement and double fault are increased. Table 5.3 reports the diversity measures calculated between ES and EM based-OHSIS systems.

Table 5.3: Diversity measures between ES and EM based-OHSIS systems.

Diversity measure method	Number of reference signatures			
	2	3	4	5
Q-statistic	**0.7855**	0.8352	0.8411	0.8847
Correlation	**0.0011**	0.0016	0.0019	0.0026
Disagreement	**0.2033**	0.1568	0.1312	0.1049
Double fault	**0.0980**	0.0657	0.0487	0.0428

For all diversity methods, the obtained results indicate that the ES and EM based-OHSIS systems are more diverse for a small number of reference signatures. This fact is highlighted by the obtained diversity scores, where Q-statistic and correlation provide the least values and the disagreement while double fault provide the highest values for 2 reference signatures. In this case, the combination is more suitable for small number of reference signatures. Considering these results, the following section reports the design of the combined individual OHSIS by means of the Choquet fuzzy integral for a possible improvement of the writer identification.

5.4.4.2. Choquet fuzzy integral design for OHSIS

The combination performed by means of the Choquet FI operator requires tuning the δ parameter defined in the fuzzy densities (Eq. 5.6). To deal with that, 10 writers (2 signatures for training and 22 for finding the best parameters) selected for designing the individual OHSIS are also used for finding the most suitable δ parameter value. Figure 7 depicts the IR of the combined OHSIS with various values of the calibration parameter δ ranging from 0.05 to 1 with a step of 0.05. It can be clearly noticed that the performance of the combined OHSIS is highly affected by the δ parameter since the IR varies from 93.18% to 97.72% when varying this parameter between [0.05, 1]. This effect justifies its introduction to the proposed density measure function.

As it is depicted in figure 5.9, the optimal calibration parameter value is fixed to $\delta_{opt} = 0.15$ to achieve the best IR for the combined OHSIS and all writers during the design step. This value will be also kept during the evaluation step when adding new writers.

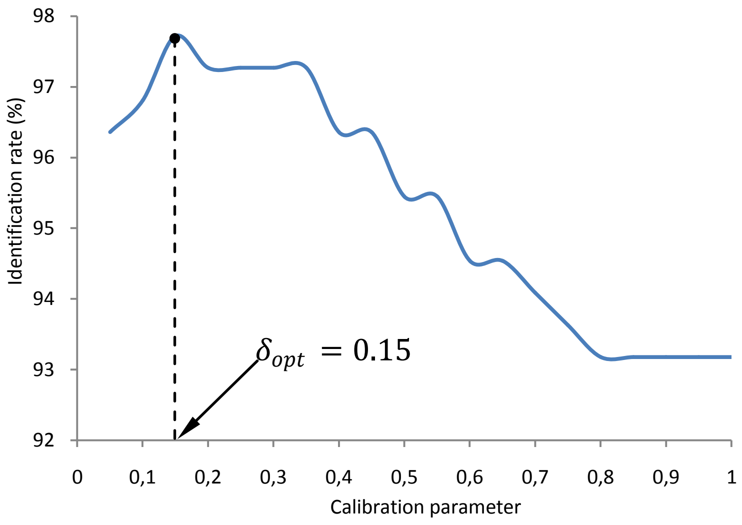


Figure 5.9: Effect of tuning the calibration parameter δ.

5.4.4.3. Combined OHSIS results and analysis

For a fair comparison of the proposed combination based on Choquet FI against other combination rules, obtained results are compared against the most used fixed rules as the Mean and Product combination rules (Kuncheva, 2004). Figure 5.10 reports the IR of the individual and combined OHSIS when writers are added progressively to the system. As can be seen, the Mean and Product combination rules achieve roughly similar results to the OHSIS based on EM descriptor. Therefore, these rules cannot take advantage of the additional information provided by the ES for OHSIS. In contrast, the results achieved by the Choquet FI show an important improvement of the IR along the 300 used writers. Moreover, the Choquet FI is the most stable combination rule when adding new writers. Indeed, Choquet FI provides the smallest IR difference when extending the system from 10 to 300 writers, which reveals more robust when extending the system for new writers.

For a thorough analysis, the OHSIS is evaluated with different numbers of reference signatures as reported in table 5.4. The Choquet FI clearly outperforms the Mean, Product and best individual OHSISs for the different number of reference signatures. More precisely, the IR is enhanced when using 2 reference signatures than 5 ones. Indeed, the achieved improvement is 4% and 2% when using 2 and 5 reference signatures, respectively. Thus, previous finding in table 5.3 confirms that the IR is improved when the diversity is more important. Consequently, the combination of different OHSIS is more suitable for small number of reference signatures.

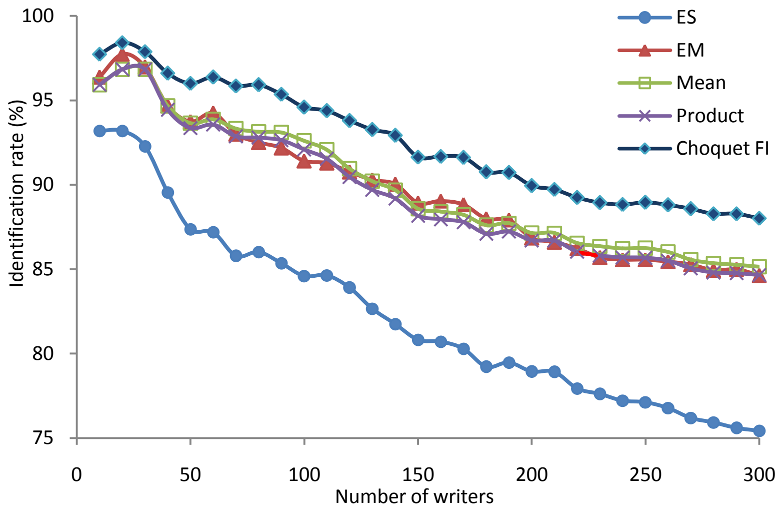


Figure 5.10: IR of individual and combined OHSIS for progressively increased number of writers trained with 2 reference signatures per writer.

Table 5.4: Writer IR (%) of individual and combined OHSIS for various reference signature numbers performed on GPDS dataset.

OHSIS	Number of reference signatures			
	2	3	4	5
ES	75.43	81.63	85.65	88.19
EM	84.62	89.53	91.50	92.75
Mean	85.13	89.66	92.11	93.45
Product	84.65	89.28	91.91	93.26
Choquet FI	**88.01**	**91.88**	**93.83**	**94.96**

To prove again the effective use of the Choquet FI combination against the Mean and Product combination rules, the McNemar's test (Dietterich, 1998) is used for comparing statistically two systems. More precisely, a contingency table is constructed in order to calculate the p-value (Dietterich, 1998). It can provide whether one system is significantly better than another according to the p-value. More the *p*-value is smaller, more the difference of accuracy is likely to be more significant. However, when the *p*-value exceeds 0.05, then the hypothesis is considered as null. In this case, both systems perform closely and the difference is too small to decide the superiority of one system than the other. Table 5.5 reports the *p*-value of the Choquet FI against the Mean and Product combination rules for various numbers of reference signatures.

The obtained results show that the *p*-values are very small than 0.05 and therefore the proposed Choquet FI is significantly better than other combination rules. Besides, the improvement is more significant when the number of reference signatures is small.

The success of Choquet FI is due to its ability to capture interactions among the OCCs and attribute the right importance for individual OHSIS. Moreover, the proposed method for fuzzy measure, which is associated with FI provides a dynamic measure allowing to give more importance to the relevant system relatively to the others for each test signature.

Table 5.5: The p-values of McNemar's test for Choquet FI versus Mean and Product combination rules.

Choquet versus	Number of reference signatures			
	2	3	4	5
Mean	$< 10^{-16}$	$1.89\ 10^{-16}$	$1.23\ 10^{-11}$	$1.83\ 10^{-10}$
Product	$< 10^{-16}$	$< 10^{-16}$	$2.53\ 10^{-13}$	$4.81\ 10^{-12}$

5.4.5. Evaluation on blind dataset

In order to show the behavior of the parameter-independent OHSIS performed on another dataset, CEDAR is used as a blind dataset in which the optimal parameters are the same as found when designing the OHSIS with GPDS. In other words, the CEDAR dataset is assumed as an extension for the GPDS dataset and therefore the same parameters are used.

Table 5.6 reports the IR obtained when testing on CEDAR dataset containing 55 writers and different number of reference signatures. It can be easily concluded that Choquet FI outperforms the other systems based on the Mean, Product and best individual OHSISs for various numbers of reference signatures and, therefore, confirms the results when the proposed OHSIS combination system is performed on GPDS dataset. Besides, the IR is more enhanced when using 2 reference signatures than 5 ones.

Also, these results exhibit the effectiveness of the designing protocol that has been proposed to find the optimal parameters. Indeed, the parameters found when designing GPDS dataset have been successfully used for CEDAR dataset and results are highly enhanced.

Table 5 6: Writer IR (%) of individual and combined OHSIS for various numbers of the reference signatures performed on CEDAR dataset.

OHSIS	Number of reference signatures			
	2	3	4	5
ES	84.04	87.87	91.36	92.05
EM	94.38	96.01	97.00	97.51
Mean	89.66	92.55	94.63	95.02
Prod	88.76	91.94	94.27	95.02
Choquet FI	**95.45**	**97.22**	**97.63**	**97.99**

5.4.6. Comparative analysis

In order to compare against the existing identification methods, few works have been found in the literature and most of them have used private datasets. Thus, the comparison of the proposed OHSIS against the state-of-the-art systems is not trivial. For instance, the evaluation of the Bi-SVM against MLP has been performed on a private dataset [9]. So far, no evaluation has been reported in the literature for GPDS dataset for the identification problem. Regarding the CEDAR dataset, only one work has been reported in literature (Kalera et al., 2004). Thus, for a fair comparison, the (Martinez et al., 2006) approach has been replicated in this work for testing on GPDS and CEDAR datasets. In addition, the comparison of the proposed parameter-independent OHSIS is performed against the Bi-SVM used by Martinez et al., (2006), where better results have been obtained than the MLP classifier on private dataset containing 38 writers. Table 5.7 reports the IR (%) achieved by the proposed individual and combined systems against other systems, when using GPDS dataset.

As it can be clearly viewed from the reported results in table 5.7, the proposed parameter-independent OHSIS highly outperforms approach (Martinez et al., 2006) for both CEDAR and GPDS datasets as well as for private dataset. Also for CEDAR dataset, the proposed system outperforms the found results reported in (Kalera et al., 2004). Indeed, the obtained IR is 93.33% using 16 reference signatures for training, and the proposed system achieves 95.45% using only 2 reference signatures. Furthermore, when examining carefully the obtained results, two observations can be deduced. First, the feature generation based on the CT is more efficient than the Bitmap using the same Bi-SVM classifier. The inefficiency of Martinez et al. approach is related to the used feature vector, which does not capture more precisely the local information contained into a signature. Indeed, the resize of each image into a fixed size 24×48 will cause a degradation of the signature shape. Second, the OC-PCA outperforms clearly the Bi-SVM classifier for both feature generation methods specifically for few reference signatures. Moreover, the proposed system based on OC-PCA requires less computational cost and can be easily deployed than the Bi-SVM since the OHSIS is able to be extended to new writers without retraining the overall system. This advantage highlights the OC-PCA as more suitable classifier to the offline handwritten signature identification problem.

Comparatively to the state-of-art, the main findings of this work are two folds. First, the CT generates an efficient feature vector for offline handwritten signature characterization. Moreover, the PCA offers an open system and absorb the high feature vector size of CT to generate a good representative model, which performs Bi-SVM classifier in terms of identification rate and training computational cost. In addition, a combination scheme based on FI is presented, which offers improved results with a small number of reference signatures. Finally, the proposed designing protocol based on a limited number of writers and reference signatures shows its effectiveness to find the optimal parameters, which have been used to extend new writers on GPDS dataset and also on CEDAR as blind dataset.

Table 5.7: IR (%) achieved by the proposed OHSIS comparatively to other state-of-the-art systems. ES, EM, CT and FI are the acronyms of Equi-Space, Equi-Mass, Curvelet Transform and Fuzzy Integral, respectively.

Reference	Dataset	#Writers	Features	Classifier	#Signatures	IR (%)
Martinez et al., 2006	Private	38	Bitmap	Bi-SVM	1	71.20
				MLP	1	46.80
Martinez et al., 2006	GPDS	300	Bitmap	Bi-SVM	5	46.31
Proposed I			ES-CT	Bi-SVM	2	**74.95**
				OC-PCA	2	**75.43**
Proposed II			EM-CT	Bi-SVM	2	**83.86**
				OC-PCA	2	**84.62**
Proposed combined			ES-CT and EM-CT	OC-PCA-FI	5	**94.96**
					4	93.83
					3	91.88
					2	88.01
Martinez et al., 2006	CEDAR	55	Bitmap	Bi-SVM	5	62.20
Kalera et al., 2004			Quasi-multiresolution	K-NN	16	93.33
Proposed combined			ES-CT and EM-CT	OC-PCA-FI	5	**97.99**
					4	**97.63**
					3	**97.22**
					2	**95.45**

5.5. Conclusion

In this chapter we have explored the hybrid OCC ensemble for the signature identification application. Thus, we have proposed a new efficient OHSIS using conjointly the curvelet transform, the OC-PCA classifiers and the Choquet fuzzy integral as a combiner for improving the writer identification specifically when the number of reference signatures is small. The main advantage of the proposed system is the ability of the curvelet transform to capture the local features contained into the signature image in order to differentiate between writers. In order to achieve an open system, new writers can be added to an existing system without training everything anew. The OC-PCA is selected to solve the identification problem since it has the advantage to absorb the high feature size and allows generating a robust representative model for each writer based on few reference signatures. Furthermore, a combination scheme based on Choquet FI of different individual HSISs is proposed by means of a new dynamic density measure for each questioned signature. Besides, an alternative protocol is proposed to design a parameter-independent open HSIS.

The experiments conducted on the well-known GPDS and CEDAR datasets show that the proposed OHSIS is able to cope with a restricted number of writers and reference signatures to design an efficient and parameter-independent OHSIS. More specifically, the OC-PCA showed very competitive results against the well know Bi-SVM classifier. Indeed, with few reference signatures, it is easy to fine-tune each writer separately by using OC-PCA instead separating only one writer from the remaining ones when using Bi-SVM. Furthermore, the combined system offers a better improvement and stability for extending new writers. In addition, with very strict design conditions, the obtained results outperform the state-of-the-art systems and confirm the validity of the proposed OHSIS, which offers a practical solution for the fundamental problem of the handwritten signature identification with a limited number of reference signatures.

For the continuation of this chapter, a new combination architecture is presented based on the hybrid OCC ensembles for the text-independent writer identification, which is more complex and interesting application.

Chapter 6

Two Stage Combination of OCC ensembles for Writer Identification of offline Handwritten Fragments

Abstract

Writer identification based on handwritten fragments has been reported to give interesting performance. However, while the fragmentation process, inconsistent fragments are generated and affect badly the identification accuracy. Hence, in this chapter, a clustered-based One-Class Classifier (OCC) is proposed in order to generate more robust classification model than the distance-based classifier for handwritten fragments. Besides, the problem of inconsistent fragments expands its effect to the test step. Thus, a Dynamic Fragment Weighting Combination (DFWC) rule is proposed to reduce the effect of inconsistent test fragments. Furthermore, due to the difficulty of performing a generic descriptor, three different descriptors related systems are designed and combined through an effective combination scheme based on Choquet fuzzy integral operator. Experimental results conducted on the well-known IFN/ENIT and IAM datasets show good adaptation of the OCC with DFWC. Moreover, the Choquet combination scheme offers more improvements to achieve 97.56% and 94.51% for the used datasets, respectively. The obtained results highlight the reliability of the proposed system in comparison with recent studies for writer identification issue.

6.1. Introduction

Person authentication based on the handwriting style is one of the oldest biometric methods. Although a huge effort has been done for proposing efficient writer identification systems, it remains however one of the most challenging problems due to the large intra-writer variations, also known as natural variation in forensic literature. Besides to its crucial use in wide areas such as digital right management in the financial sphere, to solve the expert problems in criminology by forensic expert decision-making systems, in law enforcement agencies for proving someone's authenticity and helping to distinguish the authenticity of historical handwriting documents (Brink et al., 2012; Gilliam et al., 2010; Fecker et al., 2014).

Writer identification is the process of determining the author of an unknown sample of handwritten text. Hence, automatic writer identification can be categorized according to online and offline acquisition modes. Online writer identification makes use the pen movement and dynamics of the handwritten text captured by digitizing tablets or pressure sensitive pens. In contrast, the offline mode employs the handwritten text itself, which is acquired from scanning or photographing. Therefore, offline writer identification is more complex due to the absence of such stable dynamic characteristics (Nakamur and Kidode, 2015; Sreeraj et al., 2011).

The design of an offline writer identification system (OWIS) is composed of two main modules, which are feature generation and classification. Different ways have been investigated *via* performing various feature generation methods. Thus, the design of OWIS can be also categorized according to the type of input data for capturing features, which can be whole document, paragraph, line, word, character (Sreeraj et al., 2011) and recently text fragments (Tang et al., 2013 ; Hannad et al., 2016). The use of the text fragments has a better ability of characterizing the writer's style when applying textural descriptor on the small fragments. Indeed, Tang et al. (2013) have proposed the use of handwritten fragments for the reason that texture-based methods need an important number of handwriting text to get stable features. Furthermore, non-text regions generate unstable features caused by the empty background regions contained into handwriting images. Quite recently, Hannad et al. (2016) have also explored the advantage of using text fragments. The authors claim that, the local texture descriptors are more effective and have high discriminatory power when applied on small writing fragments in characterizing the writer. Furthermore, it is worth noting that, in real word application, the whole text document of a writer cannot be always available. Therefore, the design of the OWIS based on text fragments is considered as an interesting alternative way for writer identification since only a small amount of handwriting text is required.

Generally, the distance-based classifier is the most popular to solve the writer identification problem for its fast implementation, supports high number of writers and offers an open writer identification system. The distance-based classifier has been devoted for the different types of acquired data including fragment (Nakamur and Kidode, 2015; Tang et al., 2013; Hannad et al., 2016). Indeed, Tang et al. (2013) has employed the chi-square distance for the similarity measurement between the reference and test fragments. In a recent work, Hannad et al. (2016), use the Hamming distance between the feature vectors of both reference and test fragments.

The distance-based classifier relies on the assumption that reference or training fragments form distinct clusters. Nevertheless, due to the difficulty of ensuring consistent fragments for each writer, reference fragments of various writers overlap. Therefore, more adapted classifier is required for classifying more efficiently the fragments. Quite recently, considerable attention has been paid to the One Class Classifier (OCC) for resolving many multi-class classification problems (Cyganek, 2000; Goh et al., 2005; Ban and Abe, 2006; Rabaoui et al., 2008 ; Yeh et al., 2009; Boehm et al., 2011), due to its advantages against some usual implementations based on binary or multi-class classifiers. In fact, the used OCCs do not require retraining for a second time when adding new classes to the classification system. Therefore, an open multi-class classification system is offered.

Hence, this work proposes to investigate the use of OCCs to keep an open system for identifying the writer from handwriting fragments. More precisely, the selected OCCs used for designing the OWIS are based on clustering algorithms such as K-Means and k-Centers (Tax, 2001). The main assumption of using these OCCs is that the feature vectors are clustered and can be characterized by a few prototype objects in order to generate more robust classification model based on text fragments via reducing the effect of the inconsistent ones. Since, each document is divided into different fragments, the writer identification of a query document relies on combining the generated scores from fragments matched with the reference ones (Tang et al., 2013; Hannad et al., 2016). Thus, a simple average combination rule has been adopted to combine the fragment dissimilarity scores as reported in (Tang et al., 2013; Hannad et al., 2016). However, the problem of inconsistent fragment that has been faced in the training step expands its effect to the test step. Indeed, inconsistent fragments can be also generated from the query document which may badly affect a correct writer identification. Hence, in order to reduce the incorrect classification yielding from the inconsistent fragments, this work proposes a Dynamic Fragment Weighting Combination (DEWC) rule to reduce the effect of inconsistent fragments. This rule allows giving more importance to the consistent ones. In other words, a high contribution is assigned to the more consistent fragments relatively to others.

Different feature generation methods have been performed for the writer identification achieving diverse performance for the different types of input data including text fragment (Srihari, 2002; Hertel and Bunke, 2003; Tan, 1992; Marti et al., 2011). As a result, finding the best descriptor is very difficult for all classes (i.e. writers). Hence, since some descriptors are diverse, the possible way for improving the performance of the OWIS is to combine multiple systems fed by different descriptors. The multiple classifier system is adopted in the aim to deduce a better decision from multiple opinions of diverse systems. Thus, OWISs have taken a benefit for combining multiple classifiers. Indeed, various combination strategies have been proposed aiming to provide more robust OWIS than single one (Xu et al., 1992). For instance, Bulacu et al. (2007) has proposed combination scheme using different features based on the average or weighted of the individual features participating in the combination. In related work, Fornés et al. (2010) explored a combination approach to perform the OWIS in old music scores. The K-nearest neighbor classifier is applied on the different extracted features. The resulting outputs are combined for a final classification. For this purpose, the combination is performed using the majority voting or Borda count method. Abdi and Khemakhem (2012) described a combination of multiple Arabic OWISs designed by various descriptors and distance metrics such as chi-square and Euclidean distances for the classification step. Each individual system results in a rank list, where the most probable writers are assigned to the first ranks and the least probable writers are ranked the last ones. Then, the rank lists are combined using an experimentally proven weighting formula. Recently, Tang et al. (2013) explored a weighted average rule to combine two descriptors for the writer identification problem. Both normalize chi-square distances are combined to form more efficient dissimilarity measure.

Since, the combination rule is the successful key to achieve an efficient multiple classifier system, this chapter proposes an alternative way for combining different OWISs attempting to investigate their diversity to perform a robust writer identification system. Historically, Fuzzy Integral (FI) and the associated fuzzy measures have been introduced by Sugeno for its excellent performance for classifier aggregation (Kuncheva, 2004). Its main advantage is related to measuring the strength not separately for each classifier alone but for all members. For several years, great effort has been devoted to use the FI for performance enhancement produced by multiple classifier systems. Indeed, FI has been successfully explored for various applications of pattern recognition including face detection (Kwak and Pedrycz, 2006), change detection (Nemmour and Chibani, 2006), soft biometric (Nemmour, 2015) and handwritten signature identification (Hadjadji et al. 2017).

Hence, in this chapter two stages combination system based on OCC ensembles is proposed, where the first stage is devoted for fragments combination based on the proposed DFWA rule. On the other hand, the second stage is dedicated for combining different writer identification systems via the FI combination strategy.

This chapter is organized as follows. In the next section, we introduce the proposed Two Stage Combination of Weighting and Choquet Fuzzy Integral for Writer Identification of offline Handwritten Fragments. Section 3 details the experiments along with the obtained results and analysis. The last section is devoted for the conclusions and potential research directions for writer identification.

6.2. Two stage combination scheme for writer identification

This section describes the proposed writer identification system. Thus, a description of the different modules for designing an individual writer identification system is given, which are feature generation and classification. Furthermore, a design of multiple writer identification system is also highlighted based on FI combination.

6.2.1. Writing fragmentation

The text-independent writer identification relies on characterizing the writing style of each individual writer. The forms of one character written by various people are different and in most situations characters are connected together whilst writing. Thus, rather than extracting each single character which is hard to perform, we assume that, the writing of each writer is composed of specific small fragments which occur frequently.

For the fragmentation process, in this work the one proposed by (Hannad et al., 2016) is adopted. This technique is summarized in the following steps:

1- Handwritten images binarization via a global thresholding.
2- Connected components extraction from the image.
3- Each connected component is scanned in the left-right, top-bottom fashion and is divided into small windows of size $N \times N$ generating small writing fragments

The window size N is determined empirically and the impact of window size on the overall system performance is presented in (Hannad et al., 2016).

6.2.2. Feature generation

Feature generation is an important stage for any pattern recognition system. For the writer identification problem, various methods have been explored for the different types of acquired data. The texture-based approach is known as the most used and effective for the writer identification. For the writing fragment-based writer identification, the Local Binary Pattern (LBP), Local Phase Quantization (LPQ) and Local Ternary Patterns (LTP) descriptors have been compared. The LPQ shows the best adaptation to capture the writer’s style from the writing fragments. This chapter follows the successful path of using texture-based descriptors for writing fragment by exploring the Run Length (RL), oriented Basic Image Feature (oBIF) as well as the LPQ features, which has already been explored for this problem (Hannad et al., 2016).

In the following, the mathematical foundations of the RL, oBIF and LPQ descriptors for writing fragments characterization are briefly reviewed.

6.2.2.1. Run length features

The run length (RL) texture-based features (Galloway, 1975), has been initially proposed by (Arazi et al. 1977), for writer identification. RL is performed on binary image assuming that either the black pixels corresponding to the ink trace or, more beneficially, the white pixels corresponding to the background.

For writer identification, the writing style for a given writer is characterized by computing the probability distribution of RL features contained into the writing fragments. The RL is performed on detected edges by applying Sobel operator on writing fragments. Hence, four main directions are scanned for capturing the edge pixels: horizontal, vertical, left-diagonal and right-diagonal to compute run-lengths of white and black pixels. The normalized histogram of these run lengths is interpreted as a probability distribution characterizing the writer (Arazi, 1977; Bulacu et al. 2007).

The term ‘run’ has been defined as a sequence of connected pixels which have the same color along a given direction. Consequently, the borders of each run differ from the run color. In a given direction, the run-length matrix P element (i,j) is constructed, which specifies the number of times that the writing fragment contains a run of length j, consisting of ink or background indicating by i equal to “0” or “1”, respectively. Thus, the size of the matrix P is two lines representing the ink and the background while the number of column is equal to the maximum possible run-length in the corresponding writing fragment.

Finally, the four run-length P matrices are normalized into probability distribution vectors and then concatenated to obtain a global vector characterizing the writing fragment. In order to avoid the problem of large dimensionality of the feature vector and not consistent features, only a first sub-set of columns is selected for each of the matrices. Hence, the first 10 columns are selected for each of the four directions for black and white pixels. This gives a total of 80 features ($2 \times 10 \times 4$) for each writing fragment.

6.2.2.2. Oriented basic image feature

The oriented Basic Image Feature (oBIF) is a texture-based scheme that has been successfully used for character recognition (Newell and Griffin, 2011.), texture classification (Newell et al., 2010) and recently for writer identification based on writing document (Ojansivu and Heikkila, 2008). In this work, the oBIF is used to characterize the writer's style from small writing fragments.

The oBIF is an extension to the Basic Image Feature (BIF), where a combination of the local orientation with the local symmetry information is added. When applying the oBIF descriptor on the writing fragment, each location is classified according to the local symmetry type using a bank of six Derivative-of-Gaussian (DoG) filters of a size determined by the scale parameter. Hence, seven possible symmetry types are defined which are slope, dark line, light line, dark rotational, light rotational, saddle-like and flat. The classification of any location as a flat or the other symmetry types relies on an extra parameter ε.

The slope type is accompanied by assigned orientation. However, the dark line, light line and saddle types are accompanied by an unsigned orientation. These orientations are quantized and the number of possible unsigned orientations is given by the parameter ϕ. Consequently, there are 2ϕ possible orientations for the slope type and ϕ possible orientations for each of the dark line, light line and saddle types. The dark rotational, light rotational and flat types have no orientation. Thus, a total of $5\phi + 3$ different possible oBIF types are obtained. The value of ϕ is tuned for each application. However, previous work has indicated that a value of 4 is appropriate for near-optimal performance (Newell and Griffin, 2011). Using this value, 23 oBIF types are defined for generating a feature vector. In summary, the oBIF calculation is performed according to the following steps:

1. Measure filter responses c_{ij} to an (i, j) order DoG filter, and deduce the scale normalized filter responses $s_{ij} = \sigma^{i+j} c_{ij}$
2. Compute $\mu = s_{20} + s_{02}$, $\nu = \sqrt{(s_{20} - s_{02})^2 + 4s_{11}^2}$

3. Assign BIF type according to which Expression is largest, then calculate orientation appropriately:

Expression	BIF type	Quantizable orientation
εs_{00}	flat	No orientation
$2\sqrt{s_{10}^2 + s_{01}^2}$	slope	$arctan\left(\frac{s_{01}}{s_{10}}\right) \quad s_{10} > 0$
		$arctan\left(\frac{s_{01}}{s_{10}}\right) + \pi \; s_{01} \geq 0, s_{10} < 0$
		$arctan\left(\frac{s_{01}}{s_{10}}\right) - \pi \; s_{01} < 0, s_{10} < 0$
μ	Dark rotation	No orientation
$-\mu$	Light rotation	No orientation
$(\nu + \mu)/\sqrt{2}$	Dark line	$arctan \frac{2s_{11}}{(s_{02} - s_{20} + \mu)}$
$(\nu - \mu)/\sqrt{2}$	Light line	$arctan \frac{2s_{11}}{(s_{02} - s_{20} + \mu)}$
μ	Saddle-like	$arctan \frac{2s_{11}}{(s_{02} - s_{20} + \mu)}$

6.2.2.3. Local phase quantization

The LPQ (Ojansivu and Heikkila, 2008) is known as one of the most effective features for the writer identification problem (Hannad et al., 2016). The Short Term Fourier Transform (STFT) is used to deduce the local phase information, which forms the LPQ components. The calculation of the STFT is performed for each location x in the image $f(x)$ through a rectangular neighborhood N_x of $M \times M$ size as follows:

$$F(u, s) = \sum_{y \in N_x} f(x - y) e^{-j2\pi u^T y} = w_u^T f_x \tag{6.1}$$

where w_u and f_x denote the basis vector of the discrete Fourier transform at frequency u and a vector containing all M^2 image components from N_x, respectively.

In LPQ, only four complex coefficients are considered corresponding to 2-D frequencies $u_1 = [a, 0]^T, u_2 = [0, a]^T, u_3 = [a, a]^T, u_4 = [a, -a]^T$; where a is a scalar sufficiently small to satisfy $H(u_i) > 0$. Consequently, the vector $F(x)$ is formed as follows:

$$F(x) = [F(u_1, s), F(u_2, s), F(u_3, s), F(u_4, s)] \tag{6.2}$$

The phase information provided by the Fourier coefficients is recorded in $G(x)$ vector by observing the signs of the real and imaginary parts of each component in $F(x)$. Formally, $G(x)$ is represented as:

$$G(x) = [Re\{F(x)\}, Im\{F(x)\}\,] \tag{6.3}$$

Let g_j be the j^{th} component of $G(x)$, then the sign of the phase information provided by the Fourier coefficients is determined via exploring a simple scalar quantization:

$$s_j = \begin{cases} 1, & if \;\; g_j \geq 0. \\ 0, & \text{otherwise} \end{cases} \tag{6.4}$$

The quantized coefficients are represented as integer values between 0-255 using binary coding:

$$b = \textstyle\sum_{j=1}^{8} sc_j \, 2^{j-1} \tag{6.5}$$

For writer characterization, a histogram of these integer values for all pixel locations is computed for representing the feature vector. As recommended by Hannad et al. (2016), the last bin of the histogram is not taken into consideration. Thus, this non-discriminative bin is discarded. Consequently, each writing fragment is represented by a 255 bin histogram.

6.2.3. OCC for writer identification based on writing fragments

The choice of OCC instead of the usual used distance classifier is due to two reasons. First, the OCC offers an open scheme for adding new writers to the identification system, unlike some classifiers such as support vector machine and neural networks. Second, the fragmented text can contain probably inconsistent fragments. Therefore, reference fragments of various writers overlap and consequently cannot be taken as references when matching with test fragments. Hence, the one-class K-Means and k-Centers are explored in order to generate more consistent references, based on a combination of all training fragments.

Formally, the classifier is a module that has the ability to transform a feature vector into a decision score or class label. For the writer identification application, a set $C = \{c_1, \dots, c_M\}$ of M writers are considered and each one has its corresponding function f_m associated to each

OCC_m $(m = 1,..,M)$. The training fragments are represented by distinct clusters which are described by a set of prototype vectors.

After obtaining the writer models in the training step, a query document is assigned to its corresponding writer based on combination of multiple decisions generated by the test fragments. More precisely, let a query document Q represented by a set of fragments q_i, $\{i = 1,..,Card(Q)\}$ then each test fragment q_i is compared with each writer's model f_m^{OCC} for producing a score as follows:

$$score(q_i, c_m) = f_m^{OCC}(q_i) \tag{6.6}$$

A combination of these scores over all the fragments in the test writing is necessary to measure the similarity between the query and the reference writing. Several combination rules are possible to achieve the aggregation, but all these rules need a unique interpretation of the outputs generated by the OCC for each test fragment q_i. Hence, a simple exponential function is used for transforming the OCC output $f_m^{OCC}(q_i)$ ranging from the interval $]-\infty, 0]$ into posteriori probability $P\big(c_m/(q_i)\big)$ ranged in the interval $]0,\ 1]$ expressed as follows:

$$P(c_m/q_i) = exp\left(\, f_m^i(q_i)\right) \tag{6.7}$$

The sum rule has already been used for fragments combination in order to measure the posteriori probability of the query document with each writer's model, which is expressed as follows:

$$P(c_m/Q) = \frac{1}{Card\ (Q)} \sum_{i=1}^{Card\ (Q)} exp\big(f_m^{OCC}(q_i)\big) \tag{6.8}$$

Finally, the writer of query document is identified from the set of the reference base according to the class which reports the maximum posteriori probability:

$$writer(Q) = \underset{m=1}{\overset{M}{argmax}}\big(P(c_m/Q)\big) \tag{6.9}$$

6.2.4. Dynamic fragment weighting combination rule

In the proposed scheme, the writer identification of query document relies on the combination of the different existing fragments. However, due to the high writer's style variability, some passages in the document are slightly different to maintain his/her stable style. Therefore, inconsistent fragments are generated during the fragmentation process. Consequently, the problem of inconsistent fragments that has been faced in the training step is occurred during the test step as well. Indeed, inconsistent fragments may badly affect a correct writer identification. Hence, in order to decrease the incorrect identification yielding from the inconsistent fragments, this chapter proposes a weighting combination rule to reduce the effect of inconsistent fragments. This way allows giving more importance to the consistent ones.

Generally, the static weighted average (SWA) is considered as a successful straightforward combination rule, which has been widely used for combining classifiers classifiers (Verikas et al., 1999; Al-Ani and Deriche, 2002; Wang et al., 2002) as a weighted sum and weighted Borda count rules. In addition, the SWA has been also used for multimodal biometric authentication (Snelick and Uludag, 2005), where weights are assigned to each matcher for biometric fusion. In the present work, weights are assigned to each individual fragment according to its importance for characterizing the stable writer's style. However, although the great success of fixed weighting rule for classifier combination, its use is hard to be adopted for fragment combination. This is due to the fact that fragments vary in their number and consistency according to the query document. Indeed, attributing pre-defined weight values for the query fragments is difficult and hard to be performed. Besides, recently dynamic weighted average has been successfully performed for OCC combination against static weighted average and fixed rules (Hadjadji et al., 2014a).

Hence, this chapter proposes a new and more adapted dynamic fragment weighting combination (DFWC) rule, which allows calculating the appropriate weight dynamically for each test fragment. DFWC allows matching the writer m relatively to its stable writing style characterization as follows:

$$W_m(q_i) = exp\left(-\tau\left|exp\left(f_m^{OCC}(q_i)\right) - exp\left(\bar{f}_m^{OCC}\right)\right|\right) \tag{6.10}$$

Such that, $0 \leq W_m(q_i) \leq 1$ and $0 < \delta \leq 1$.

$$\bar{f}_m^{OCC} = \frac{1}{Card\,(Q_m)} \sum_{i=1}^{Card\,(Q_m)} f_m^{OCC}(r_i) \tag{6.11}$$

$\bar{f}_m^{OCC}$ is the average of all reference fragment r_i outputs contained in the training document Q_m. Therefore, $\bar{f}_m^{OCC}$ represents the writing stable style response for writer m. δ is a positive value that has been introduced for more calibrating the difference between the test fragment response and the writer's stable style. Hence, the closer this difference is, the greater the weight value is. In contrast, when this difference is important, the test fragment characterizes different writing style than the matched writer. Therefore, a small weight value is assigned. Consequently, the proposed DFWC allows generating the posteriori probability according to the following equation:

$$P(c_m/Q) = \frac{1}{\sum_{j=1}^{Card\,(Q)} W_m(q_j)} \sum_{i=1}^{Card\,(Q)} W_m(q_i) exp\big(f_m^{OCC}(q_i)\big) \tag{6.12}$$

The proposed DFWC is then performed for generating dynamically the posteriori probability instead of simple sum or average rule.

6.2.1. FI combination of multiple writer identification systems

The characterization of the writing style based on writing fragments using texture descriptors has shown good identification results. Moreover, some descriptors are diverse and it is possible to explore this property by performing a combination scheme. The combination can be performed on the feature level either on the decision or the score. The decision level is the most successful way since it is allowed to give more importance separately for each system unlike the feature level where all descriptors are gathered to form a joint feature vector.

Various combinations strategies have been developed for pattern recognition systems including writer identification. In this system, the FI is explored to combine different systems fed by three texture descriptors, which are RL, oBIF and LPQ. Thus, the proposed system is designed by using two combination stages. The first stage is devoted for combining writing fragments based on the proposed DFWC rule. While, the second stage uses the FI, which is performed for combining the obtained results by the three writer identification systems. As it is depicted in Figure. 6.1, the proposed system is composed of M writers and 3 different texture descriptors. Consequently, three different classifier models are generated per writer.

The FI operator is applied after the DFWC rule. Therefore, the Choquet operator is performed to handle the three posteriori probabilities $P_i(c_m/Q)$ yielding from the LPQ, RL and oBIF based writer identification system. The resulting decision value defines the score of the query document belonging to each writer. Hence, the evidence values for the Choquet FI are expressed as the posteriori probability taking the following form:

$$h_m(z_i) = P_i(c_m/Q) \tag{6.13}$$

The FI relies on a pre-knowledge on the different information sources to be combined. Hence, an accurate evaluation of each individual system is required by exploiting the density measures to ensure the success of the combination. Generally, a validation dataset is used for pre-evaluating the system for getting the representative density measure values. However, this approach assumes that all testing samples are represented by the same density measure values. Consequently, a less efficient representation is yielded since each query document has its appropriate density measure. Therefore, more contribution should be given in respect to the more appropriate writer identification system to the current query document.

To overcome this problem, a similar function to the proposed DFWC is proposed in this work for calculating the density measure associated to each classifier defined as follows:

$$g_m^i(Q) = exp\left(-\delta \left\| P_i(c_m/Q) - exp\left(\bar{f}_m^i\right) \right\|^2\right) \tag{6.14}$$

Such that, $0 \leq g_m^i(x) \leq 1$ and $0 < \delta \leq 1$.

Such that, $0 \leq g_m^i(x) \leq 1$ and $0 < \delta \leq 1$. τ is a positive parameter that has been involved for calibrating the density measure. This parameter is tuned till finding the most suitable value for representing the contribution of each individual writer identification system. The density measure is calculated according to the similarity between the stable writing style $\bar{f}_m^i$ and the posteriori probability $P_i(c_m/Q)$ generated from the DFWC rule. Thus, the density measure values are generated dynamically for each query document.

In this work, it is worth noting that the writing style parameter is involved for both combination stages. In the first stage, it is involved to calculate the contribution of each fragment by means of the DFWC rule to perform an individual writer identification system. While, in the second stage, this parameter is also involved to measure the strength of each individual system by means of the density measures, leading to an effective multiple writer identification system combination.

After obtaining the density measures *via* applying the proposed technique, the fuzzy measures, $g_m(A_i)$ for $1 \leq i \leq L$ and $1 \leq m \leq M$, are calculated by applying Eqs. 1.24 and 1.25.

Let Ic_m as the Choquet aggregation produced from the posteriori probabilities of the LPQ, RL and oBIF writer identification systems associated to their corresponding fuzzy measures defined for each writer, then the query document Q is assigned to the writer class with the maximal value as follows:

$$Writer(Q) = \underset{m=1}{\overset{M}{argmax}}\,(Ic_m(Q)) \tag{6.15}$$

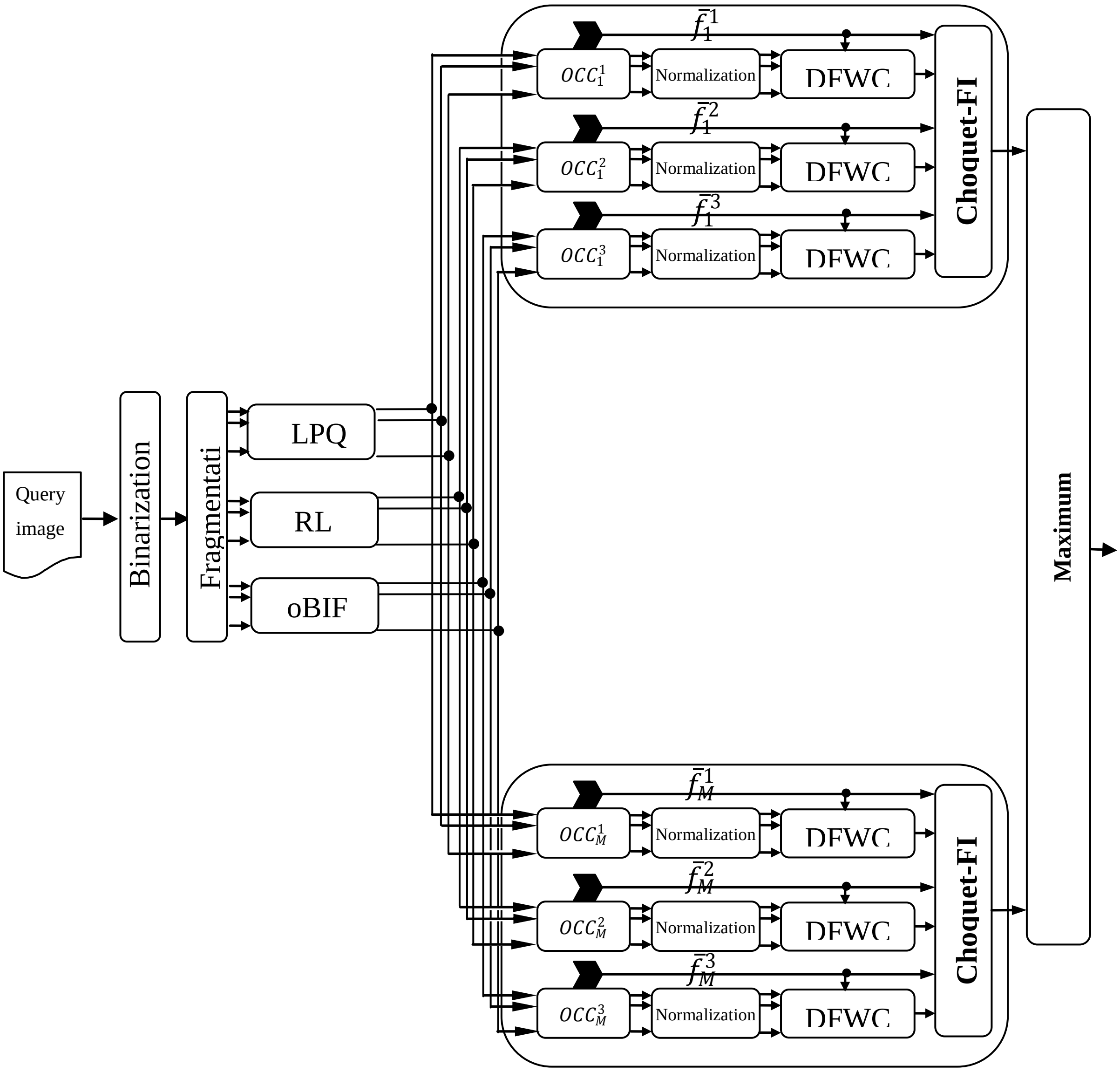


Figure 6.1: Two stage combination writer identification system

In summary, the proposed writer identification system is performed according to the Algorithm 6.1.

Algorithm 6.1:

Inputs: Query fragments q, classifier models OCC_j^i, stable writing style $\bar{f}_m^i$

Output: Writer identification

1: for $i \leftarrow 1$ to M do /* i represents class variable */

2: for $j \leftarrow 1$ to L do /* j represents the feature generation method variable */

3: for f ← 1 to $Card(Q)$ do /* f represents the query fragment variable */

4: Calculate the fragment feature vector x using j^{th} method.

5: Calculate the classifier output according to Eq. 6.6

6: Compute the normalized output $P_i\left(c_j/x\right)$ according to Eq. 6.7

7: Compute the weight value of the current fragment according to Eq. 6.10

8: end for

9: Apply the DFWC to aggregate the query fragments results according to Eq. 6.12

10: Compute the evidence value $h_j(z_i)$ according to Eq. 6.13

11: Calculate the density measure according to Eq. 6.14

12: end for

13: Determine the fuzzy measures $g_j(A_i)$, for $1 \leq i \leq L$ and $1 \leq j \leq m$ using Eqs. 1.24 and 1.25.

14: Perform the Choquet FI integral according to Eq. 1.28

15: end for

16: Assign the unknown written fragments to the writer model that provides the maximum FI combination score according to Eq. 6.15.

6.3. Experimental results

6.3.1. Databases and design protocol

The performance of the proposed writer identification system is evaluated on two popular datasets namely IFN/ENIT and IAM. The well known IFN/ENIT dataset (www.ifnenit.com) is a collection of Arabic handwritten documents containing more than 26400 images of Tunisian town names written in Arabic script. Words are written by 411 writers using different writing tools, which make it more challenging. While the IAM dataset (Marti and Bunke, 2002) is a collection of 1,539 forms contributed by 657 different writers containing 13,353 isolated and labeled handwritten English text lines of variable content.

For generating the text fragments, the one proposed by Hannad et al. (2016) is adopted in this work. This technique is summarized in the following steps:

- Binarization of handwritten images via a global thresholding.
- Extraction of connected components from the image.
- Generation of small writing fragments by dividing each connected component into a window size $N \times N$. The window size is determined empirically and its impact on the overall system performance has been evaluated in (Hannad et al., 2016).

For both datasets, the same fragment fractions for training and testing are used in this chapter as reported in (Hannad et al., 2016) for far comparison. According to Hannad et al. (2016), for each of the 411 writers of the IFN/ENIT dataset, 30 word images are used to create the reference document R and 20 word images per writer forms the query document Q. While, for each of 657 writers of the IAM dataset, the same training and testing fragments are used. In this case, 60% of the text lines are used as the reference base while the remaining 40% are used for evaluation.

The performance of the proposed system is evaluated using the Identification Rate (IR) expressed in %, which is defined as the number of instances correctly identified to the total number of instances formally formulated as:

$$\mathrm{IR} = 100 \times \frac{\text{Number of correct identification for all writers}}{\text{Total number of test documents for all writers}} \ (\%) \tag{6.15}$$

Furthermore, for performing an optimal design of the writer identification system, various parameters should be tuned, which are related to the OCCs (the number of cluster K and center k), DFWC (δ) and the density measures (τ) for performing Choquet operator. In this work, the training dataset is divided into two subsets. The first one composed of 2/3 training dataset is used to train the system with different values. While the second remaining 1/3 dataset is used for finding the optimal parameters to achieve the best IR. Once optimal parameters are found, the OCCs are retrained again on the whole training dataset in order to get more stable models.

6.3.2. Evaluation of OCCs for WI using handwritten fragments

The individual OWI systems are evaluated on both K-Means and k-Centers OCCs for different handwritten fragments. Thus, Tables 6.1 and 6.2 report comparative results for

OCCs against distance dissimilarity based-classifier (Tang et al., 2013; Hannad et al., 2016), for the different descriptors conducted on IFN/ENIT and IAM datasets, respectively.

The obtained results clearly show that the used descriptor affects highly the identification rate for both datasets. Indeed, the LPQ and RL provide roughly similar results in contrast to the oBIF descriptor which provides lower results. Furthermore, the OCC based on K-Means performs better than the OCC based on k-Centers for all descriptors and both dataset. However, the distance dissimilarity based- classifier outperforms the best OCC in some cases, for instance, when using LPQ in both datasets. In contrast, surprising results are achieved by the oBIF descriptor for IAM dataset. Hence, the OCC is reliable to maintain stable results for the different descriptors as well as its adaptation for the different datasets in comparison to the distance classifier. However, the lower IR achieved by the OCCs in other cases are explained by its sensitivity to the presence of inconsistent fragments in the query document. These fragments are considered as outliers which may badly affect the correct writer identification (Tax, 2001). In order to take into consideration this fact, the next section presents the DFWC having the ability to reduce the effect of the inconsistent fragments and to provide more robust writer identification system.

Table 6.1: Writer IR (%) for OCCs against distance dissimilarity based-classifier performed on IFN/ENIT database.

Descriptor	Classifier		
	K-means	k-center	Distance
LPQ	93.67	88.32	**94.40**
Run length	**93.18**	88.07	92.45
oBIF	82.23	78.34	**90.51**

Table 6.2: Writer IR (%) for OCCs against distance dissimilarity based-classifier performed on IAM database

Descriptor	Classifier		
	K-means	k-center	Distance
LPQ	83.23	71.34	**88.26**
Run length	84.75	66.92	80.64
oBIF	68.44	60.06	**29.72**

6.3.3. Dynamic fragment weighting versus average combination

In this section, the proposed DFWC rule is evaluated against the already used average fragment combination rule (Tang et al., 2013; Hannad et al., 2016), which is considered as Static Fragment Weighting Combination (SFWC) since weights are the same and fixed for all fragments. Thus, Tables 6.3 and 6.4 report comparative results obtained when performing both fragment–combination methods on the used classifiers and the different descriptors conducted on IFN/ENIT and IAM datasets, respectively.

From the obtained results, the DFWC shows its effectiveness to improve widely the IR against the average or SFWC rule for the OCCs. For instance, when using oBIF descriptor with one class k-Centers classifier, the IR is improved from 60.06% to 84.29% for IAM dataset. Moreover, after performing the DFWC on OCCs based clustering methods, the K-Means classifier outperforms the most widely distance dissimilarity based classifier for both LPQ and RL descriptors. However, when using the distance dissimilarity based classifier, the IR is not improved a lot when performing the DFWC rule. The inappropriate adaptation of the DFWC for the distance dissimilarity based classifier is due to the presence of inconsistent fragments in the reference dataset, unlike the OCCs where new reference dataset is formed. In fact, the presence of inconsistent fragments in the reference dataset does not allow differentiating between consistent and inconsistent test fragments. This outcome justifies the choice of using the OCC based on clustering algorithm for the writer identification based on handwritten fragments.

Furthermore, it is worth noting that the RL descriptor with one-class K-Means classifier form the best individual writer identification system by achieving 96.35% and 92.22% for IFN/ENIT and IAM datasets, respectively.

Table 6.3: Writer IR (%) for DFWC against Average combination performed on IFN/ENIT database

Descriptor	Classifier					
	K-means		k-center		Distance	
	Average	DFWC	Average	DFWC	Average	DFWC
LPQ	93.67	**95.86**	88.32	94.16	94.40	94.40
Run length	93.18	**96.35**	88.07	93.67	92.45	93.18
oBIF	82.23	88.56	78.34	87.10	**90.51**	90.51

Table 6.4: Writer IR (%) for DFWC against Average combination performed on IAM database

Descriptor	Classifier					
	K-means		k-center		Distance	
	Average	DFWC	Average	DFWC	Average	DFWC
LPQ	83.23	91.00	71.34	**91.61**	88.26	89.17
Run length	84.75	**92.22**	66.92	89.48	80.64	80.48
oBIF	68.44	84.14	60.06	**84.29**	29.72	32.01

6.3.4. Choquet FI for multiple writer identification systems

This section is devoted for presenting the IR obtained for the second combination stage of the proposed system. As the first combination stage is performed by the DFWC to achieve individual system, the second combination stage consists of combining the different individual writer identification systems leading to take benefit from their diverse descriptors. For a fairly comparison of the proposed Choquet FI combination against other aggregators, the IRs are compared against the standard used fixed rules (Kuncheva, 2004) as the Average, Product, Maximum (Max) and Minimum (Min).

Thus, for each type of the used classifier, the individual systems associated to the respective LPQ, RL and oBIF descriptor are combined via fixed rules and the Choquet operator. Tables 6.5 and 6.6 report the obtained IRs for IFN/ENIT and IAM dataset, respectively.

Table 6.5: Writer IR (%) of two stages combined system for the different classifiers performed on IFN/ENIT database.

Combination rule	Classifier					
	K-means		k-center		Distance	
	Average	DFWC	Average	DFWC	Average	DFWC
average	90.99	95.37	91.24	**96.35**	94.40	94.40
product	89.78	93.67	89.29	**95.62**	94.89	94.89
max	93.67	**95.86**	88.07	94.16	94.40	94.40
min	82.23	88.56	78.34	87.10	**90.51**	**90.51**

Choquet-FI	94.64	**97.56**	94.40	97.08	96.10	96.10

Table 6.6: Writer IR (%) of two stages combined system for the different classifiers performed on IAM database.

Combination rule	Classifier					
	K-means		k-center		Distance	
	Average	DFWC	Average	DFWC	Average	DFWC
average	73.62	90.85	67.98	**93.75**	83.35	83.53
product	83.53	89.63	71.64	**92.83**	80.33	81.40
max	83.23	91.00	71.34	**91.61**	88.26	89.17
min	68.44	84.14	60.06	**84.29**	29.72	**32.01**
Choquet-FI	79.72	**94.20**	69.20	90.85	90.09	89.32

The first glance of the obtained results shows that the combination of the different individual systems performed by the DFWC is more suitable since it provides improved IRs for the different combinations of the second stage. Secondly, when observing the combined writer identification systems, it can noted that the Choquet combiner performs an improved IR against fixed combiners for the different types of classifier. The good performance of Choquet FI is owing to its capability to capture diversity between individual systems and attribute the right importance for each one. Besides, the proposed method for density measure values, which are generated dynamically for each query document, allows providing more importance to the relevant system relatively to the others.

The simultaneous use of the K-Means classifier, DFWC and Choquet constitutes the best choice for providing the best performance for both datasets. Indeed, while using K-Means classifier, the proposed combinations in both stages allow improving the identification rate from 93.67 to 97.56 and also from 83.23 to 94.51 for IFN/ENIT and IAM datasets, respectively. Finally, the OC based on the K-Means is more suitable for the writer identification based on writing fragments than the usual distance classifier. This outcome is due to its propriety to reduce the presence of inconsistent fragments, which offers a better adaptation with DFWC and FI unlike the distance classifier.

6.3.5. Stability study of combined systems

In order to study the performance stability of the proposed system according to the number of writers, a series of experiments were carried out by adding progressively new of writers. Figures 6.2 and 6.3 depicts the identification rate according to the number of writers achieved by the Choquet, average, product, max and min combinations to perform the multiple writer identification systems for IFN/ENIT and IAM databases, respectively. We can clearly notice an expected drop in the identification rate when adding progressively new writers. More precisely, the Choquet shows its efficiency to keep a stable performance according to the number of writers unlike the average, product, and max which show a drop in performance after exceeding 50 writers. However, the min rule achieves the worst results. Consequently, the Choquet operator provides the best performance and the least affected when adding new writers. Indeed, Choquet based identification system yields the smallest IR difference when adding new writers for both databases.

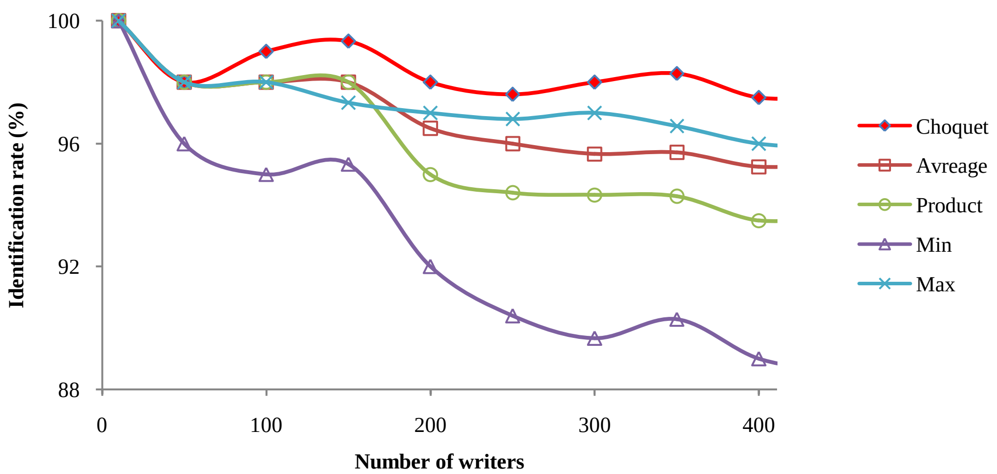


Figure 6.2: Identification rate on IFN/ENIT database in function with number of writers

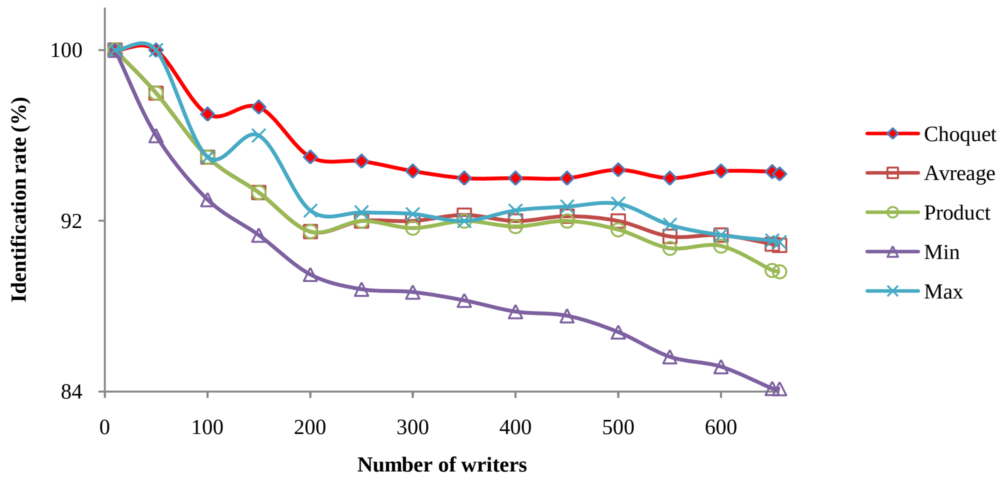


Figure 6.3: Identification rate on IFN/ENIT database in function with number of writers

In addition, the stability of the best proposed tow stages combinations system based on K-means classifier is evaluated for reduced number of training fragments. In fact, in real environment, each writer will be willing to provide a small amount of handwritten and therefore it is important to study its effect on the system performances. Thus, we design the system with a reduced number of fragments for training and we keep the overall test dataset. Figures 6.4 and 6.5 depict the IR of the different combinations in function with training dataset percentage for IFN/ENIT and IAM databases, respectively. From the obtained results, a natural increase of the IR when an important number of fragments is used for the training is observed for both databases. Furthermore, the Choquet reveals to be the best combiner and the least affected for the different percentages of the training set. Indeed, for IFN/ENIT database while the Choquet based system drops by around 17% of IR when decreasing the training set to 40%, the best fixed combiner drops by around 23%. Consequently, the combination of different writer identification systems is more suitable for reduced number of training fragments.

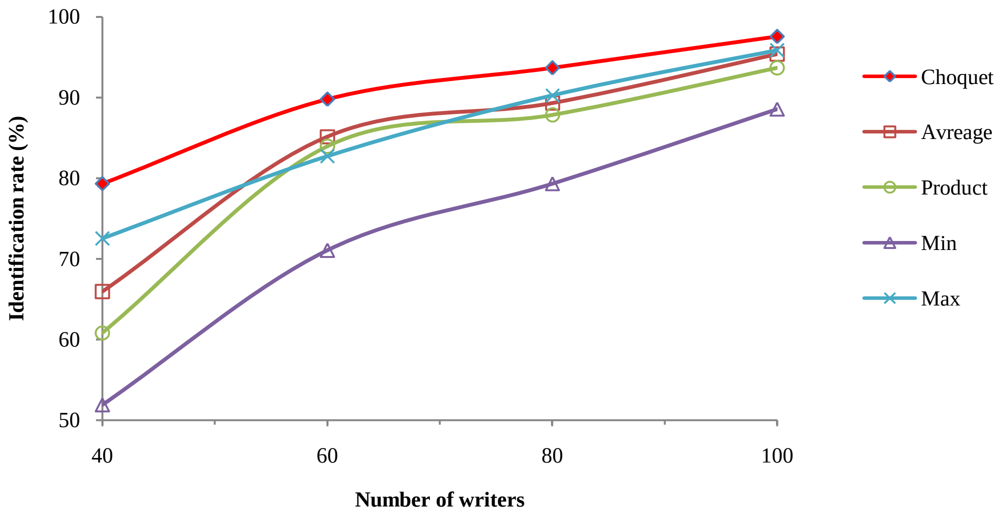


Figure 6.4: Identification rate on IFN/ENIT database in function with training set percentage

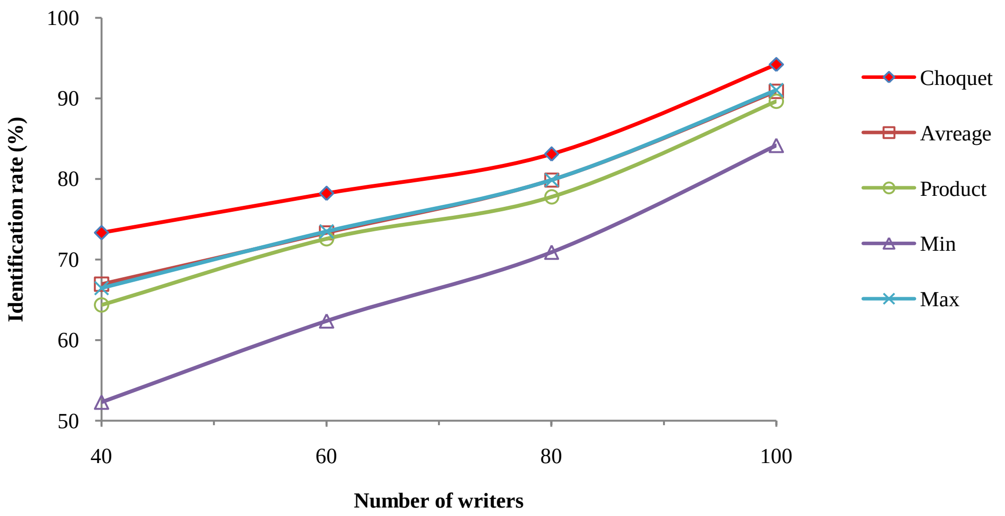


Figure 6.5: Identification rate on IAM database in function with training set percentage

6.3.6. Comparative analysis

In order to situate the proposed system among the state-of-the-art systems, a comparative study of identification performance is highlighted for a number of studies available in the literature. Only studies have evaluated IFN/ENIT and IAM datasets are presented in Tables 6.7 and 6.8, respectively. For IFN/ENIT dataset, many systems have been proposed leading to provide an efficient Arabic writer identification system. Indeed, different types of input data have been performed. For instance, Khan et al. (2016) have used the whole document to achieve 76%. Bulacu et al. (2007) have used a paragraph as input data to achieve 88% on 350 writers. In other work, Abdi and Khemakhem (2015) have achieved around 90% on 411 writers by using text line as input data. Recently, results of Arabic script have been widely improved by Hannad et al. (2016) by achieving 94.90% on 411 writers using written fragments. However, the proposed system offers more improvements via achieving 97.56% for the IFN/ENIT dataset via using writing fragments as input data.

Moreover, for IAM dataset, various systems have been developed as shown in Table 8. From the presented results, the most powerful systems use the document as input data. Indeed, Bertolini et al. (2013), Wu et al. (2014) and Khan, et al. (2016) have achieved excellent results for the writer identification. However, these systems require a whole document for an accurate identification, which is hard to obtain in real environment. Thus, in a recent work, Hannad et al. (2016) have proposed the use of handwriting fragments for the writer identification, 89% have been obtained on 657 writers of IAM dataset.

Table 6.7: Writer IR (%) achieved by the proposed system comparatively to other state-of-the-art systems for IFN/ENIT database

Type of data	System	# writer	feature	classifier	combination	IR (%)
Document	Djeddi and Souici-Meslati (2010)	130	RL	Euclidian	--	82.00
	Djeddi and Souici-Meslati (2011)	130	Co-occurrence features	Artificial immune recognition	--	84.23
	Khan, et al. (2016)	411	Discrete cousin transform	Kernel discriminant analysis with spectral regression	Majority voting	76.00
Paragraph	Bulacu et al. (2007)	350	directional, grapheme and run-length	$\chi 2$	Weighted average	88.00
Line	Abdi and Khemakhem (2015)	411	Elliptic graphemes	$\chi 2$	--	90.02
Fragment	Hanned et al. 2016	411	LPQ	Hamming	--	94.90
	Our approach	411	LPQ, RL and oBIF	OC-K-means	Chqouet FI	**97.56**

Nevertheless, the proposed two stage combination system achieves 94.51% on the same dataset, which is very challenging and comparable to the state-of-the-art. Indeed, a reduced set of samples per writer are used in order to simulate real world conditions.

Besides, the obtained results reveal the good flexibility of the proposed system to handle two different languages Arabic and English, which is generally hard to perform.

Table 6.8: Writer IR (%) achieved by the proposed system comparatively to other state-of-the-art systems for IAM database

Type of data	Study	# writer	Feature	classifier	combination	IR (%)
Document	Siddiqi and Vincent (2010)	650	Codebook and Contour	$\chi 2$	Average	89.00
	Bertolini et al. (2013)	650	LPQ	Support vector machine	--	96.70
	Wu et al. (2014)	657	scale invariant feature transform	$\chi 2$ and Euclidien	Weighted sum	**98.50**
	He and Schomaker (2014)	650	Delta–n Hinge features	$\chi 2$	--	93.20
	He et al. (2015)	650	junction feature	Kohonen 2D	--	91.10
	Khalifaa et al. (2015)	650	grapheme codebook	Kernel discriminant analysis with spectral regression	--	92.00
	He and Schomaker (2017)	650	RL of LBP and Cloud Of Line Distribution	$\chi 2$	Weighted average	89.90
	Khan et al. (2016)	650	Discrete cousin transform	Kernel discriminant analysis with spectral regression	Majority voting	97.20
Paragraph	Bulacu and Schomaker (2007)	650	directional, grapheme and RL	$\chi 2$	Weighted average	89.00
Fragment	Hanned et al. 2016	657	LPQ	Hamming	--	89.11
	Proposed	657	LPQ, RL and oBIF	OC-K-means	Chqouet FI	94.20

6.4. Conclusion

This chapter aimed to propose a writer identification system using handwritten fragments. Hence, a new two stage combination of OCC is proposed based on DFWC and Choquet FI techniques. Due to the high writer's style variation, some generated fragments are inconsistent and do not represent the authentic writing style. Indeed, the correct identification is highly affected by the inconsistent fragments when using the distance-based classifier. Thus, this work has explored the clustered-based OCC for their capability to reconstruct new more representative reference fragments based on a combination of all training fragments. Furthermore, the problem of inconsistent fragments expands its affect to the test step since inconsistent fragments are also contained in the query documents. Also, the DFWC is proposed in this work in order to give more importance to the fragments having the ability to characterize well the stable writing style of each writer. Moreover, the selection of a good descriptor is very hard to deal with fragment approach. Hence, the improvement of the identification is possible by exploiting the existence of diverse descriptors via combining multiple systems fed by different descriptors. Thus, in order to reach a good decision from multiple combinations of various systems, a fuzzy integral combination is proposed in this work to build a multiple Arabic writer identification system. In particular, three different descriptors related systems are designed and combined through an effective combination scheme based on Choquet operator.

The experiments conducted on the well-known IFN/ENIT and IAM datasets show good results of the proposed system. First, the clustered-based OCC shows great adaptation with DFWC unlike the distance classifier. The no adaptation of the DFWC for the distance dissimilarity based classifier is due to the presence of inconsistent reference fragments in the matching dataset, unlike the OCCs where new references are formed. In fact, the presence of inconsistent fragments in the matching dataset does not allow differentiating between consistent and inconsistent test fragments. This behavior justifies the choice to use the OCC based on clustering algorithm for the writer identification using handwritten fragments. Furthermore, when applying the DFWC to perform individual writer identification systems, the experimental results show that the cluster-based OCCs outperform the distance-based classifier. More precisely, the RL descriptor with the OC-K-Means classifier provides best results for both datasets. Besides, the combination of different descriptors by Choquet combiner performs an improved recognition rate against fixed combiners for the different types of classifier. However, the K-Means is observed to provide the best results for both datasets. Furthermore, the obtained results outperform the state-of-the-art systems and confirm the validity of the proposed system, which offers a practical solution for the fundamental problem of writer identification using handwritten fragments.

As a future work, a new architecture is planned to propose a selection of fragments for training and testing for a better reduction of the inconsistent fragment effect.

Conclusion

Automatic pattern classification is related to various real-world applications such as the classification of a given email to "spam" or "non-spam". Moreover, the classification problem becomes more complex when the number of classes is increased such as in the application of person identification based on biometric information (Schwartz *et al*., 2012). The classifier bloc allows to convert an input feature vector into class vector and therefore attribute it to one of a given classes. Various classification methods have been proposed and investigated in the literature according to the aimed application. However, although the good offered classification accuracy, these techniques perform closed system. Indeed, usual multi-class classifiers take in consideration either the entire classes for generating the classification model, such as the neural networks, or by splitting up the original problem into a set of two-class sub-problems. Nowadays, extended multi-class implementation to new classes is strongly required, for many interesting applications. However, the existing classifiers need to retrain the system again on all classes such as the OAO or OAA implementations based on the SVM classifiers. Recently, OCC has been successfully used to achieve extensible multi-class implementations. Indeed, extending the OCC to new classes does not require retraining the used classifiers for a second time. Furthermore, the OCC offers less computational cost in terms of training time and memory space. However, using a single system of OCC for the multi-class implementation usually achieves less accuracy than the usual multi-class implementations.

Thus, this thesis proposed to explore multiple multi-class implementation systems based on OCC leading to perform higher accuracy and keep at the same time the open classification system. More precisely, the objective of the presented thesis is the enhancement of OCC ensembles for solving the multi-class classification problem. The proposed contributions have been devoted for real world applications in general and specifically for some handwritten recognition applications such as word recognition, signature identification and writer identification.

Contributions: From the research that was carried out throughout this thesis, five main contributions were reported as summarized as follows:

1- Combination of different types of OCC for multi-class classification by means a new Dynamic Weighted Average (DWA) combination rule. Experimental results

2- conducted on several real-world datasets prove the effective use of the proposed approach where the DWA rule achieves the best results against fixed rules as well as the decision template. Furthermore, comparison of the proposed open classification system against a standard open classifier based on K-Nearest Neighbor shows the superiority of the proposed system.

3- An improved one class Auto Associative Neural Network ensemble is proposed based on a selection algorithm for selecting the appropriate training samples to the AANN, which allows enhancing the AANN combination and classification robustness. Experimental results conducted on several real-world datasets prove the effective use of the proposed algorithm to construct more robust AANN ensembles.

4- A combination scheme of OCCs is proposed based on fuzzy integral operators. Furthermore, an alternative framework is proposed to design a parameter-independent and open-lexicon handwritten Arabic word recognition system as well as a new density measure function. Experimental results conducted on Arabic handwritten dataset using different types of OCCs with large number of classes show the superiority of FI for OCC ensembles

5- An Open Handwritten Signature Identification System is proposed by using conjointly the Curvelet Transform and the One-Class classifier based on Principal Component Analysis. A combination based on Choquet fuzzy integral is explored to combine multiple individual OHSISs. Experimental results conducted on standard CEDAR and GPDS handwritten signature datasets report 97.99% and 94.96% correct identification rate, respectively, which highlights the effectiveness of the proposed OHSIS since it can comfortably outperform the state-of-the-art when using few reference signatures.

6- Two stage combination system is proposed for writer identification based on handwriting fragments. The first stage is devoted for combining fragments based on a proposed dynamic fragment weighting combination rule. On the other hand, the second stage is dedicated for combining different writer identification systems fed by three descriptors via the FI combination strategy. Experimental results conducted on the well-known IFN/ENIT and IAM databases show good adaptation of the OCC with DFWC. Moreover, the Choquet combination scheme offers more improvements to achieve 97.56% and 94.20% for the used databases, respectively. The obtained results highlight the reliability of the proposed system in comparison with recent studies for writer identification issue.

Future works: The future prospects of this research consist to develop other combination rules such as the one based on neutrosophic probability which has been recently presented. Furthermore, it is interesting to evaluate the proposed combination schemes on other applications of multi-class classification, such as multi-script writer identification and word spotting for historical documents.

References

Abbas, N., Chibani, Y., Belhadi, Z., and Hedir, M., 2013, A DSmT based combination scheme for multiclass classification, in: 16th International Conference on Information Fusion, Istanbul, Turkey, pp. 1950-1957.

Abdi M. N., Khemakhem, M., 2015, A model-based approach to offline text-independent Arabic writer identification and verification, Pattern Recognition, 48(5) 1890–1903.

Abdi M.N., Khemakhem, M., 2012, Arabic Writer Identification and Verification Using Template Matching Analysis of Texture, 2012 IEEE 12th International Conference on Computer and Information Technology (CIT), Chengdu, China , pp. 592 – 597 , 2012.

Al-Ani, A., Deriche, M., 2002, A new technique for combining multiple classifiers using the Dempster-Shafer theory of evidence, Journal of Artificial Intelligence Research 17 333-361.

Alginahi, Y. M., 2013, A survey on Arabic character segmentation, International Journal on Document Analysis and Recognition (IJDAR) 16(2) 105-126.

AlKhateeb, J.H., Pauplin, O., Ren, J., Jiang, J., 2011, Performance of hidden Markov model and dynamic Bayesian network classifiers on handwritten Arabic word recognition, Knowledge-Based Systems 24 680–688.

Arazi, B., 1977, Handwriting identification by means of run-length measurements, In IEEE Transactions on systems, man and cybernetics, 7(12) 878 - 881.

Arivazhagan, S., Ganesan, L., Kumar, T. G. S., 2006, Texture classification using Curvelet statistical and co-occurrence features, in: International Conference on Pattern Recognition, Hong Kong, pp. 938-941.

Ban, T., Abe, S., 2006, Implementing multi-class classifiers by one-class classification methods, In: International Joint Conference on Neural Networks, pp. 327-332, Vancouver, BC, Canada.

Batista, L., Granger, E., Sabourin, R., 2012, Dynamic selection of generative-discriminative ensembles for off-line signature verification, Pattern recognition 45 1326-1340.

Bergani, C., Oliveira, L.S., Koreich, A.L., Sabourin, R., 2009, Combining different biometric traits with one-class classification, Signal Processing (89) 2117- 2127.

Bertolini, D., Oliveira, L., Justiona, E., and Sabourin, R., 2010, Reducing forgeries in writer independent off-line signature verification through ensemble of classifiers, Pattern Recognition (43) 387-396.

Bertolini, D., Oliveira, L.S., Justino E.J.R., Sabourin, R., 2013, Texture-based descriptors for writer identification and verification, Expert Systems with Application, 40(6) 2069–2080.

Bishop, C. M., 2006, Pattern Recognition and Machine Learning. Information Science and Statistics Series Editor, Springer.

Bishop, C.M., 1995, Neural networks for pattern recognition, Oxford University Press, Oxford.

Boehm, O., Hardoon, D.R., Manevitz, LM., 2011, Classifying cognitive states of brain activity via one-class neural networks with feature selection by genetic algorithms, International Journal of Machine Learning and Cybernetics (2) 125–134.

Boyer, K.W., Govindaraju, V. and Ratha, N.K., 2007, Special issue on recent advances in biometric systems, IEEE Transaction on Systems, Man and Cybernetics, Part B Cybernetics 37 (5) 1091-1095.

Breiman, L., 1996, Bagging predictors. Machine Learning 24, 123–140.

Breve, F.A., Ponti-Jr, M. P., and Mascarenhas, N. D. A., 2007, Multilayer perceptron classifier combination for identification of materials on noisy soil science multispectral images, in Proc. 20th Brazilian Symposium on Computer Graphics and Image Processing (SIBGRAPI 2007). Belo Horizonte, MG: IEEE, pp. 239–244.

Brink, A., Smit, J., Bulacu, M., Schomaker, L., 2012, Writer Identification Using Directional Ink-Trace Width Measurements, Pattern Recognition 45 (1) 162-171.

Bulacu, M., Schomaker, L., 2007, Text-independent writer identification and verification using textural and allographic features ,IEEE Trans.Pattern Anal .Mach. Intell. 29(4) 701–717.

Bulacu, M., Schomaker, L., Brink, A., 2007, Text-independent writer identification and verification on offline arabic handwriting. In ninth International Conference on Document Analysis and Recognition. ICDAR, pp. 769–773.

Cabrera, J.B.D., Gutiérrez, C., Mehra, R.K., 2008, Ensemble methods for anomaly detection and distributed intrusion detection in Mobile Ad-Hoc Networks, Information Fusion 9 96-119.

Candès, E. and Donoho, D., 1999, Curvelets – a surprisingly effective nonadaptive representation for objects with edges. In: A. Cohen, C. Rabut and L. Schumaker, Editors, Curves and Surface Fitting: Saint-Malo, Vanderbilt University Press, Nashville, pp. 105-120.

Candès, E., Demanet, L., Donoho, D. L., and Ying, L., 2006, Fast discrete curvelet transforms, Multiscale Modeling and Simulation (5) 861–899.

Cheedella, S. Gupta, S. R. Mahadeva Prasanna, and B. Yegnanarayana, 2002, Autoassociative neural network models for online speaker verification using source features from vowels, INNS-IEEE International Joint Conference on Neural Networks, Hawaii, USA, May 12-17

Cheplygina, V., Tax, D. M. J., 1999, Pruned Random Subspace Method for One-Class Classifiers. In: Multiple Classifier Systems. Lecture Notes in Computer Science, vol. 6713, pp. 96–105.

Chiang, J.H., 1997, Gaber, P.D., Hybrid fuzzy-neural systems in handwritten word recognition. IEEE Transactions on Fuzzy Systems (5) 497-510.

Chiang, J.H., 1999, Choquet fuzzy integral-based hierarchical networks for decision analysis. IEEE Transactions on Fuzzy Systems, (7) 63–71.

Cho, S.-B. 1995, Fuzzy aggregation of modular neural networks with ordered weighted averaging operators. International Journal of Approximate Reasoning. 13(4) 359–375.

Cho, S.-B., Kim, J. H., 1995, Combining multiple neural networks by fuzzy integrals for robust classification. IEEE Transactions on System, Man, and Cybernetics 25(2) 380-384.

Cottrell, G.W., Munro, P.W., Zipser, D., 1988, Image compression by back propagation: a demonstration of extensional programming. In: Sharkey NE (ed) Advances in cognitive science, vol 2. Abbex, Norwood

Cyganek, B., 2012, One-Class Support Vector Ensembles for Image Segmentation and Classification, Journal of Mathematical Imaging and Vision 42(2-3): 103-117.

Cyganek, B., Krawczyk, B., 2015, Data Classification with Ensembles of One-Class Support Vector Machines and Sparse Nonnegative Matrix Factorization, in: 7th Asian Conference on Intelligent Information and Database Systems, Bali, Indonesia, pp. 526-535.

Dietterich, T. G., 2000, Ensemble methods in machine learning, in 1st Int. Work. on Multiple Classifier Systems (MCS 2000), LNCS 1857. Cagliary, Italy: Springer, 2000, pp. 1–15.

Dietterich, T. G., Bakiri, G., 1995, Solving multiclass learning problems via error-correcting output codes, Journal of Artificial Intelligence Research (2) 263–286.

Dietterich, T.G., 1998, Approximate statistical tests for combining supervised classification learning algorithms, Neural Computation 10 (7) 1895-1923.

Dietterich, T.G., 2000, Ensemble methods in machine learning, in: Multiple Classifier Systems, eds. J. Kittler and F. Roli, vol. 1857, pp. 1–15. New York, NY, USA: Springer.

Djeddi, C., Souici-Meslati, L., 2010, A texture based approach for arabic writer identification and verification, in: Machine and Web Intelligence,(ICMWI), International Conference on, IEEE, pp. 115–120. Djeddi and Souici-Meslati.

Djeddi, C., Souici-Meslati, L., 2011, Artificial immune recognition system for arabic writer identification, in: Innovation in Information & Communication Technology (ISIICT), Fourth International Symposium on,IEEE, 2011, pp. 159–165. Djeddi and Souici-Meslati.

Duda, R. and Hart, P., 1973, Pattern Classification and Scene Analysis. John Wiley & Sons, New York.

Duda, R.O., Hart P.E., 2001, Stork, D.G., Pattern classification, 2nd ed. John Wiley &Sons, NY (2001).

Duin, R. P. W., 2002, The combining classifier: to train or not to train?, in: 16th International Conference on Pattern RecognitionCanada, pp. 765–770.

Elleuch, M., Maalej, R., Kherallah, M., 2016, A New Design Based-SVM of the CNN Classifier Architecture with Dropout for Offline Arabic Handwritten Recognition, International Conference on Computational Science Vol. 80, 2016, Pages 1712–1723, ICCS

Elrajubi, O.M., El-Feghi, I.S., 2015, Angle Features Extraction of Handwritten Signatures, in: International Conference on Computer Vision and Image Analysis Applications (ICCVIA), Tunisia, 2015

Farah N., Souici L., and Sellami M., 2005, Classifiers combination and syntax analysis for Arabic literal amount recognition," Engineering Applications of Artificial intelligence 19(1) 29-39.

Favata, J. and Srikantan, G., 1996, A Multiple Feature/ Resolution Approach To Handprinted Digit and Character Recognition, International journal of imaging systems and technology 7(4) 304–311.

Fecker, D., Asit, A., Margner, V., El-Sana, J., Fingscheidt, T., 2014, Writer Identification for Historical Arabic Documents, in: Pattern Recognition (ICPR), 22nd International Conference on, Stockholm, pp. 3050-3055.

Feng, Z., Jiang, W., He and MA, B., 2011, A recognition and novelty detection approach based on Curvelet transform, nonlinear PCA and SVM with application to indicator diagram diagnosis, Expert Systems with Applications 38 (10) 12721-12729.

Fornes, A., Llados, J., Sanchez, G., Otazu, X., Bunke, H., 2010, A combination of features for symbol-independent writer identification in old music scores. International Journal on Document Analysis and Recognition 13: 243–259.

Fumera, G. and Roli, F., 2005, A theoretical and experimental analysis of linear combiners for multiple classifier systems," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 27, pp. 924–956.

Galloway, M.M., 1975 Texture analysis using gray level run lengths, In: Computer Graphics Image Processing, 4(2) 172 – 179.

Geman, S., Bienenstock, E., and Doursat, R., 1992, Neural networks and the bias/variance dilemma, Neural Computation, vol. 4, no. 1, pp. 1–58.

Gesù, V. Di and Bosco, G. Lo., 2007, Combining one class fuzzy knn's. In Proceedings of the 7th international workshop on Fuzzy Logic and Applications: Applications of Fuzzy Sets Theory, Berlin, Heidelberg, pp. 152–160.

Giacinto, G., Perdisci, R., Rio, M. D. & Roli, F., 2006, Intrusion detection in computer networks by a modular ensemble of one-class classifiers. Information Fusion 9 69–82.

Gilliam, T., Wilson, R., Clark, J., 2010, Scribe Identification in Medieval English Manuscripts, in: Pattern Recognition (ICPR), 20th .International Conference on, Istanbul, pp. 1880-1883.

Goh K.S., Chang, E.Y., Li, B., 2005, Using one-class and two-class SVMs for multiclass image annotation, IEEE Transactions on Knowledge and Data Engineering 17 (10) 1333–1346.

Guerbai, Y., Chibani, Y., and Hadjadji, B., 2015, The effective use of the one-class SVM classifier for handwritten signature verification based on writer-independent parameters, Pattern Recognition 48(1) 103-113.

Hadjadji, B., Chibani, Y., 2014b, Optimized Selection of Training Samples for One-Class Neural Network Classifier, In; International Joint Conference on Neural Networks (IJCNN), Beijing, China, July 6 – 11, 2014, 345-349.

Hadjadji, B., Chibani, Y., 2015, Outlier's Effect Reduction of One-Class Neural Network Classifier, In: 3rd International Conference on Control, Engineering & Information Technology, Tlemcen, Algeria. May 25-27.

Hadjadji, B., Chibani, Y., Guerbai, Y., 2014a, Multiple One-Class Classifier Combination for Multi-Class Classification, In; the 22end International Conference on Pattern Recognition (ICPR), Stockholm, Sweden,August 24 – 28, 2832-2837.

Hadjadji, B., Chibani, Y., Guerbai, Y., 2017a, Combining diverse one-class classifiers by means of dynamic weighted average for multi-class pattern classification, Intelligent Data Analysis 21 (3) 515-535.

Hadjadji, B., Chibani, Y., Nemmour, H., 2014c, Fuzzy Integral Combination of One-Class Classifiers Designed for Multi-Class Classification, In; ICIAR 2014, Portugal, October 22 – 24, 320-328.

Hadjadji, B., Chibani, Y., Nemmour, H., 2017b, An Efficient Open System for Offline Handwritten Signature Identification based on Curvelet Transform and One-Class Principal Component Analysis, Neurocomputin (265) 66–77.

Han, K., Sethi, K., 1996, Handwritten signature retrieval and identification, Pattern Recognition Letters (17) 83-90.

Hannad, Y., , Siddiqi, I., , El Kettani, M.E.Y., 2016, Writer identification using texture descriptors of handwritten fragments, Expert Syst. Appl. 47 14–22

Hastie, T., Tibshirani, R., Friedman, J.H., 2001, The Elements of Statistical Learning. Springer-Verlag.

He, S., Schomaker, L., 2014, Delta-nhinge: rotation-invariant features for writer identification, in: International Conferenceon Pattern Recognition, 2014, pp. 2023–2028

He, S., Schomaker, L., 2017, Writer identification using curvature-free features, Pattern Recognition (63) 451-464

He, S., Wiering, M., Schomaker, L., 2015, Junction detection in handwritten documents and its application to writer identification, Pattern Recognition, 48(12), pp. 4036-4048

Hertel, C., Bunke, H., 2003, A Set of Novel Features for Writer Identification,‖ Proc. Fourth International Conference Audio and Video-Based Biometric Person Authentication, 2003, pp. 679-687.

Hmeidi, I., Hawashin, B., El-Qawasmeh, E., 2008, Performance of KNN and SVM classifiers on full word Arabic articles, Advanced Engineering Informatics 22 106–111

Hoffmann, H., 2007, Kernel PCA for novelty detection”,. Pattern Recognition, 40(3):863–874.

Hsu, C.W., Lin, C.J., 2002, A comparison of methods for multiclass support vector machines, IEEE Trans. Neural Networks, 13 415-425.

Ismail, M.A., Gad, S., 2000, Off-line Arabic signature recognition and verification, Pattern Recognition 33 1727-1740.

Jain, K., Ross, A. and Prabhakar, S., 2004, An introduction to biometric recognition, IEEE Transactions on Circuits and Systems for Video Technology, Special Issue on Image- and Video-Based Biometrics, 14 (1) 4-20.

Japkowicz, N., Myers, C., and Gluck M. A.: 1995, ‘A Novelty Detection Approach to Classification’. In: Proceedings of the 14th International Conference on Artificial Intelligence (IJCAI-95). pp. 518–523.

Jayech, K., Mahjoub, M.A., Ben Amara, N.E., 2016, Synchronous Multi-Stream Hidden Markov Model for Offline Arabic Handwriting Recognition without Explicit Segmentation, Neurocomputing 214(19) 958-971.

Juszczak, P. and Duin, R.P.W., 2004, Combining one-class classifiers to classify missing data, in: 5th International Workshop, Multiple Classifier Systems, pp. 92-101. Cagliari, Italy.

Kalera, M.K., Srihariy, S., and Xu, A., 2004, Offline signature verification and identification using distances statistics, International Journal of Pattern Recognition and Artificial Intelligence, (18) 1339-1360.

Kazemi, F. M., Izadian, J., Moravejian, R., and Kazemi, E.M., 2008, Numeral recognition using curvelet transform, in: IEEE/ACS International Conference on computer Systems and Applications, Doha, Qatar, pp. 606-612.

Khalifa, E., Al-maadeed, S., Tahir, M.A., Bouridane, A., Jamshed, A., 2015, Off-line Writer IdentiÞcation Using an Ensemble of Grapheme Codebook Features, Pattern Recognition Letters 2015, doi: 10.1016/j.patrec.2015.03.004.

Khan, F. A., Atif Tahir, M., Khelifa, F. A., Bouridane, Resheed Almotaeryi, 2016, Robust Off-line Text Independent Writer IdentiÞcation Using Bagged Discrete Cosine Transform Features, Expert Systems With Applications (2016), doi: 10.1016/j.eswa.2016.11.012.

Kishore, S. P., Yegnanarayana, B. Suryakanth, and Gangas hetty, V., 2001, Online text-independent speaker verification system using autoassociative neural network models”. Proc. INNS-IEEE International Joint Conference on Neural Networks, Washington DC,USA, vol.2(4), pp. 1548-1553, July 14-19.

Kittler, J., Hatef M., Duin, R.P.W., Matas, J., 1998, On combining classifiers, IEEE Trans. on Pattern Analysis and Machine Intelligence (20) 226–239.

Knerr S., Augustin E., Baret O., Price D., 1998, Hidden Markov model based word recognition and its application to legal amount reading on French checks,” Journal of computer vision and image understanding 70(3) 404-419.

Kohavi, R., 1995, A study of cross-validation and bootstrap for accuracy estimation and model selection. Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence, 1995, San Mateo, CA: Morgan Kaufmann. 2 (12) 1137-1143.

Krawczyk B., Filipczuk, P., 2014, Cytological image analysis with firefly nuclei detection and hybrid one-class classification decomposition, Engineering Applications of Artificial Intelligence (31) 126-135.

Krawczyk, B., Wozniak, M., 2014, Diversity measures for one-class classifier ensembles, Neurocomputing 126, 36-44.

Krawczyk, B., Wozniak, M., 2014, Influence of Distance Measures on the Effectiveness of One-Class Classification Ensembles, Applied Artificial Intelligence 28(3): 258-271

Krawczyk, B., Woźniak, M., Cyganek, B., 2014 Clustering-based ensembles for one-class classification. Information Sciences 264, 182–195.

Kumar, R., Sharma, J.D. and Chanda, B., 2012, Writer-independent off-line signature verification using surroundedness feature, Pattern Recognition Letters 33: 301-308.

Kuncheva, L., 2004, combining pattern classifiers :methods and algorithms, Wiley-Interscience publication, New Jersey, USA.

Kwak, K-C. Pedrycz,W., 2005, Face recognition: A study in information fusion using fuzzy integral, Pattern Recognition Letters 26, Issue 26 (2005) 719–733.

Kwang-Kyu, S., 2007, An application of one-class support vector machines in content-based image retrieval, Expert System with Applications (33) (2) 491-498.

Lai, C., Tax, D.M.J., Duin, R.P.W., Pekalska, E. and Paclik, P., 2002, On combining one-class classifiers for image database retrieval, in Proceedings of Multiple Classifier Systems (Springer, Berlin, 2002), pp. 212–221.

Lawgali, A., 2015, A survey on Arabic character recognition, International Journal of Signal Processing, Image Processing and Pattern Recognition 8(2) 401-426.

Lee, Z.-Y., Yeh, C-Y, Lee, S.-J, 2007, Applying discriminant functions with one-class SVMs for multi-class classification. In: Sixth International Conference on Machine Learning and Cybernetics, Hong Kong, pp. 19-22.

Leena, M., and Yegnanarayana, B.., 2004, “Autoassociative neural network models for language Identification”. In 2004. Proceedings of International Conference, Intelligent Sensing and Information Processing, pp. 317-320, 2004.

Li, Y., Yang, Q. and Jiao, R., 2009, Image compression scheme based on curvelet transform and support vector machine, Expert Systems with Applications 37(4) (2009) 3063-3069.

Likforman-Sulem, L., Sigelle, M., 2008, Recognition of degraded characters using dynamic Bayesian networks, Pattern Recognition 41(10) 3092-3103.

Majumdar, A., 2006, Bangla basic character recognition using digital Curvelet transform, Journal of Pattern Recognition Research 2(1) 17–26.

Majumdar, A., 2009, Image compression by sparse PCA coding in curvelet domain, in: Signal Image and Video Processing. 3(1) (2009) 27-34.

Mandal, T. and Wu, Q. M. J., 2008, Face recognition using curvelet based PCA, in: International Conference on Pattern Recognition, Florida, USA, pp. 1-4.

Manevitz, L., and Yousef, M., 2007, One-class document classification via Neural Networks, Neurocomputing, vol. 70, pp. 1466–1481.

Marchette, D., 1999, A statistical method for profiling network traffic, in: Workshop on Intrusion Detection and Network Monitoring, pp. 119–128. USENIX Association. Marchette,

Marti, U.V., Bunke, H., 2002, The iam-database: an english sentence database for offline handwriting recognition. International Journal on Document Analysis and Recognition, 5(1) (2002) 39–46. Marti and Bunke.

Marti, U.V., Messerli, R., Bunke, H., 2011, Writer Identification Using Text Line Based Features,‖ Proc. Sixth International Conference on Document Analysis and Recognition (ICDAR), pp. 101-105.

Martinez, E.F., Sanchez, A., Velez, J., 2006, Support vector machines versus multi-layer perceptrons for efficient off-line signature identification, Engineering Applications of Artificial Intelligence (19) 693-704.

Moshtaghi, M., Havens, T.C., Bezdek, JC., Park, L., Leckie, C., Rajasegarar, S., Keller, J.M., Palaniswami, M., 2011, Clustering ellipses for anomaly detection, Pattern Recognition 44 (1) 55–69.

Muñoz-Marí, J., Camps-Valls, G., Gómez-Chova, L., and Calpe-Maravilla, J., 2007, Combination of one class remote sensing image classifiers, in: International Geoscience and Remote Sensing Symposium, pp. 1509-1512, 2007.

Munroe, D.T., Madden, M.G., 2005, Multi-class and single-class classification approaches to vehicle model recognition from images, in: Irish Conference on Artificial Intelligence and Cognitive Science, Portstewart.

Nakamur, Y.., Kidode, M., 2015, Individuality analysis of online kanji handwriting,‖ in International Conference on Document Analysis and Recognition Nakamur and Kidode, 2015.

Nanni, L., 2006 Experimental comparison of one-class classifiers for online signature verification, Neurocomputing 69 (7–9) 869–873.

Newell, A.J., Griffin, L., 2011, Natural image character recognition using oriented basic image features, in: Proceedings of the International Conference on Digital Image Computing: Techniques and Applications (DICTA).

Newell, A.J., Griffin, L.D., Morgan, R.M., Bull, P.A., 2010, Texture-base destimation of physical characteristics of sandgrains, in: Proceedings of the International Conference on Digital Image Computing: Techniques and Applications (DICTA), pp.504–509.

Ojansivu, V., Heikkila, J., 2008, Blur insensitive texture classification using local phase quantization. In Image and signal processing (pp. 236–243).Springer, 2008.

Palanivel, S., Venkatesh, B.S., and Yegnanarayana, B., 2003, Real time face authentication system using autoassociative neural network models”. In ICME 03, 2003 International Conference, Multimedia and Expo, volume 1, page 257.

Parzen, E., 1962, On estimation of a probability density function and mode. Annals of Mathenatical Statistics, 33:1065–1076.

Pavlidis, N.P., 1998, Papanikolopoulos, R., Mavuduru, R., Signature identification through the use of deformable structures, On Aligning Curves 71(2) 187–201.

Perdisci, R., Gu, G., Lee, W., 2006, Using an ensemble of one-class SVM classifiers to harden payload-based anomaly detection systems, in: 16th International Conference on Data Mining, Hong Kong, China, pp. 488-498.

Perdisci, R., Gu, G., Lee, W., 2006, Using an ensemble of one-class SVM classifiers to harden payload-based anomaly detection systems. In: 6th IEEE International Conference on Data Mining, pp. 488–498. IEEE Press, New York.

Pham, T., Wagner, M., 2000, Similarity normalization for speaker verification by fuzzy fusion. Pattern Recognition 33 309-315.

Pirlo, G., 1994, Algorithms for Signature verification, in: Proceedings NATO-ASI Series Fundamentals in Handwriting Recognition, S. Impedovo, Ed.Berlin, Germany: Springer-Verlag, pp. 433–454.

Plamondon, R. and Lorette, G., 1989, Automatic signature verification writer identification-The state of the art, Pattern Recognition 22(2) 107-131.

Plötz T., and Fink G.A., 2009, Markov models for offline handwriting recognition : A survey,” International Journal of Document Analysis and Recognition 12(4) 269-298.

Rabaoui, A., Davy, M., Rossignol, S., Ellouze, N., 2008, Using one-class SVMs and wavelets for audio surveillance, IEEE Trans. on Information Forensics and Security 3 (4) 763-775.

Rath, T. M. and Manmatha, R., 2003, Features for word spotting in historical manuscripts, In: 7th Int'l Conf. on Document Analysis and Recognition (Edinburgh, Scotland, August 3-6, vol. 1, pp. 218–222.

Riba, J.R., Carnicer, A., Vallmitjana, S., and Juvells, I., 2000, Methods for invariant signature classification. in: Proceedings 15th international conference on pattern recognition, pp. 2953-2956.

Ridler, T., Calvard, S., 1978, Picture thresholding using an iterative selection method, IEEE Trans. on Systems, Man and Cybernetics 8 (8) 630–632.

Sachs, A., Thiel, C., and Schwenker, F., 2006, One-class support-vector machines for the classification of bioacoustic time series. International Journal on Artificial Intelligence and Machine Learning 6 (4) 29–34.

Saha, A., and Wu, Q. J., 2010, Facial expression recognition using curvelet based local binary patterns. in: IEEE International Conference on Acoustics Speech and Signal Processing, Texas, USA, pp. 2470 – 2473.

Schölkopf, B., Platt, J.C., Shawe-Taylor, J., Smola, A.J., and Williamson, R.C., 2001, Estimating the support of a high-dimensional distribution. Neural Computation, Vol. 13, no.7, pp.1443-14472.

Schwartz W. R., Guo, H., Choi, J., Davis, L. S., 2012, Face Identification Using Large Feature Sets, IEEE Transactions on Image Processing, VOL. 21, NO. 4, APRIL, 2245-2255

Segui, S., Igual, L., Vitrià, J., 2010, Weighted bagging for graph based one-class classifiers. in: 9th International Workshop on Multiple Classifier Systems, Lecture Notes in Computer Science, vol. 5997, pp. 1–10. Springer, Heidelberg.

Seokho, K. , Sungzoon, C., 2015, A novel multi-class classification algorithm based on one-class support vector machine, Intelligent Data Analysis 19 (4) 713-725.

Shatnawi, M., 2015, Off-line handwritten Arabic character recognition: a survey, Proceedings of the International Conference on Image Processing, Computer Vision, and Pattern Recognition (IPCV), 2015.

Shieh, A. D., Kamm, D. F., 2009, Ensembles of One Class Support Vector Machines, , in: MCS '09: Proceedings of the 8th International Workshop on Multiple Classifier Systems, Reykjavik, Iceland, , pp. 181–190.

Shipp, C., Kuncheva, L., 2002, Relationships between combination methods and measures of diversity in combining classifiers, Information Fusion (3) 135-148.

Shirdhonkar, M. S., Kokare, M., 2011, Off-line handwritten signature retrieval using Curvelet transforms. International Journal of Computer and Engineering 3 (4) 1658-1665.

Shyu, M.L., Chen, S.C., Sarinnapakorn, K., Chang L., 2003, A novel anomaly detection scheme based on principal component classifier. In: IEEE Foundations and New Directions of Data Mining Workshop, pp. 172-179.

Siddiqi, I., Vincent, N., 2010, Text independent writer recognition using redundant writing patterns with contour-based orientation and curvature features, Pattern Recognition, 43(11) 3853–3865.

Spinosa, E.J., 2004, Ferreira de Carvalho, A.C. P. L., SVMs for novel class detection in bioinformatics. In: Brazilian Workshop on Bioinformatics, 2004, pp. 81–88.

Sreeraj M., Mary Idicula, S., 2011, A Survey on Writer Identification Schemes, International Journal of Computer Applications 26 (2) 0975 – 8887.

Srihari S., Cha, S., Arora, H., Lee, S., 2002, Individuality of Handwriting,‖ J. Forensic Sciences, vol. 47 (4): 1-17.

Srihari, S.N., Xu, A. and Kalera, M.K., 2004, Learning strategies and classification methods for off-line signature verification, in: Proceedings of the 9th International Workshop on Frontiers in Handwriting Recognition, pp. 161–166.

Starck, J., Candes, E. J., and Donoho, D. L., 2002, The curvelet transform for image denoising, IEEE Transactions on Image Processing 11(6) 670-684.

Sulong, G., Ebrahim, A.Y., and Jehanzeb, M., 2014, Offline Handwritten Signature Identification Using Adaptive Window Positioning Techniques, Signal & Image Processing: An International Journal (SIPIJ), 5(3) 13-24.

Sumana, I. J., Islam, M. M., Zhang, D., and Lu, G., 2008, Content based image retrieval using Curvelet transform, in the 10th Workshop on Multimedia Signal Processing, Australia, pp. 11–16.

Sun, B.Y., Huang, D.S., 2003, Support vector clustering for multi-class classification problems, in: The Congress on Evolutionary Computation, Canberra, Australia, pp. 1480–1485.

Tan, T. N., 1992, Texture feature extraction via visual cortical channel modeling,‖ in International Conference of Pattern Recognition, vol. 3, pp. 607–610.

Tang, Y., Wu, X., Bu, W., 2013, Offline text-independent writer identification using stroke fragment and contour based features, 2013 IEEE International Conference on Biometrics (ICB), pp. 1-6.

Tarassenko, L., Hayton, P., and Brady, M. ,1995, Novelty detection for the identification of masses in mammograms. In Proc. of the Fourth International IEEE Conference on Artificial Neural Networks, volume 409, pages 442–447.

Tax, D., Duin, R., 2001, Combining one-class classifiers, in: Proceedings of the Second International Workshop on Multiple Classifier Systems, MCS ’01, SpringerVerlag, London, UK, pp. 299–308.

Tax, D.M.J, 2001, One-class classification, PhD Thesis, Delft University of Technology.

Tax, D.M.J., Duin, R., 2004, Support vector data description Machine learning, Springer 54, 45-66.

Tax, D.M.J., Duin, R.P.W., 2001, Combining one-class classifiers. In: Kittler, J., Roli, F.(eds.) MCS 2001. LNCS, vol. 2096, pp. 299–308. Springer, Heidelberg.

Tax, D.M.J., Duin, R.P.W., 2005, Characterizing one-class datasets, in: Sixteenth Annual Symposium of the Pattern Recognition Association of South Africa, pp 21–26.

Tu, Y., Li, G., and Dai, H., 2006, Integrating local one-class classifiers for image retrieval, in: Advanced Data Mining and Applications, volume 4093, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 213–222.

Tumer, K. and Ghosh, J., 1996, Error correlation and error reduction in ensemble classifiers, Connection Science, vol. 8, no. 3,4, pp. 385–204.

Ullman, N., 1978, Elementary statistics, an applied approach. Wiley and Sons.

Vargas, J., Ferrer, M., Travieso, C., Alonso, J., 2007, Off-line handwritten signature GPDS-960 corpus, in: International Conference on Document Analysis and Recognition, pp 764-768.

Verikas, A., Lipnickas, A., Malmqvist, K., Bacauskiene, M., Gelzinis, A.,1999, Soft combination of neural classifiers: A comparative study, Pattern Recognition Letters (20) 429-444.

Vielhauer, C. and Dittmann, J., 2006, Biometrics for users authentification: Encyclopedia of multimedia, ed. B. Furth, Springer, Berlin, Germany.

Vinciarelli A., 2002, A survey on off-line cursive word recognition, Journal of the Pattern Recognition Society, vol.35, no.7, pp.1433-1446.

Wang, W., Brakensiek, A., Rigoll, G., 2002, Combination of Multiple Classifiers for Handwritten Word Recognition, in: International Workshop on Frontiers in Handwriting Recognition, pp. 117–122. Niagara-on-the-Lake, Canada.

Wilk, T., Wozniak, M., 2012, Soft computing methods applied to combination of one-class classifiers. Neurocomputing (75) 185-193.

Wolpert, D. H., 1996, The lack of a priori distinctions between learning algorithms," Neural Computation, vol. 8, no. 7, pp. 1341–1390.

Wozniak, M., Grana, M., Corchado, E., 2014, A survey of multiple classifier systems as hybrid systems, Information Fusion (16) 3–17.

Wu, G. and Chang, E., 2003, Adaptive Feature-Space Conformal Transformation for Learning Imbalanced Data, in: Internationl Conference on Machine Learning, Aug. 2003.

Wu, X., Tang, Y., Bu, W., 2014, Offline text-independent writer identification based on scale invariant feature transform, IEEE Trans. Inf.Forens. Secur. 9(3) 526–536.

Xu, L., Kryzak, A., Suen, C.Y., 1992, Methods of Combining Multiple Classifiers and Their Applications to Handwriting Recognition. IEEE Trans. Syst., Man, Cybern., 22(3):418–435.

Yeh, C.Y., Lee, Z.Y., Lee, S.J., 2009, Boosting one-class support vector machines for multi-class classification, Applied Artificial Intelligence 23 (4) 297 - 315.

Yong, S.P., Deng, J. and Purvis, M.K., 2012, Novelty detection in wildlife scenes through semantic context modelling, Pattern Recognition 45 (9) 3439–3450.

Zanero, S., Savaresi, S.M., 2004, Unsupervised learning techniques for an intrusion detection system, in: ACM symposium on applied computing pp. 412–419.

Zhang, Y., Zhang, B., Coenen, F., Xiao, J. and Lu, W., 2014, One-class kernel subspace ensemble for medical image classification, EURASIP Journal on Advances in Signal Processing (17) 1-13.

Scientific Production

- ## Journal publications

1. Hadjadji, B., Chibani, Y., Nemmour, H., 2017, An Efficient Open System for Offline Handwritten Signature Identification based on Curvelet Transform and One-Class Principal Component Analysis, Neurocomputin (265) 66–77. **Impact factor: 3.31**

2. **Hadjadji, B.,** Chibani, Y., Guerbai, Y., Combining diverse one-class classifiers by means of dynamic weighted average for multi-class pattern classification, Intelligent Data Analysis 21 (3) 515-535. **Impact factor: 0.772**

3. Gattal, A., Chibani, Y., **Hadjadji, B.,** 2017, Segmentation and recognition system for unknown-length handwritten digit strings, Pattern Analysis and Applications, DOI 10.1007/s10044-017-0607-x.

4. Smarandache, F., Abbas, N., Chibani, Y., **Hadjadji, B.,** Omar, Z.A., 2016, PCR5 and neutrosophic probability in target identification, Progress in Nonlinear Dynamics and Chaos, 4(2) 45-50.

5. Guerbai, Y., Chibani, Y., **Hadjadji, B.,** 2015, The effective use of the one-class SVM classifier for handwritten signature verficication based on writer-independent parametrs, Pattern Recognition 48, 103–113.

- ## Book chapter :

1. Abbas, N., Chibani, Y., **Hadjadji, B.,** 2015, A DSmT Based Combination Systems for Handwritten Signature Verification, Advances and Applications of DSmT for Information Fusion, Collected Works, Volume 4, 423- 440.

- ## International conferences

1. Abbas, N., Chibani, Y., **Hadjadji, B.,** Omar, Z.A., 2016, A DSmT based system for writer-independent handwritten signature verification, 19th International Conference on Information Fusion, Heidelberg, Germany, July 5-8, 2016.

2. Gattal, A., Chibani, Y., **Hadjadji, B**., Nemmour, H., Siddiqi, I., Djeddi, C., 2015, Segmentation-Verification Based on Fuzzy Integral for Connected Handwritten Digits Recognition, In: 5th International Conference on Image Processing Theory and applications (IPTA), Orléans, France, November 10-13, 2015.

3. **Hadjadji, B.,** Chibani, Y., 2015, Outlier's Effect Reduction of One-Class Neural Network Classifier, In: 3rd International Conference on Control, Engineering & Information Technology, Tlemcen, Algeria. May 25-27, 2015.

4. **Hadjadji, B.,** Chibani, Y., Nemmour, H., 2014, Fuzzy Integral Combination of One-Class Classifiers Designed for Multi-Class Classification, In; ICIAR 2014, Portugal, October 22 – 24, 2014, 320-328.

5. **Hadjadji, B.,** Chibani, Y., Guerbai, Y., 2014, Multiple One-Class Classifier Combination for Multi-Class Classification, In; the 22end International Conference on Pattern Recognition (ICPR), Stockholm, Sweden,August 24 – 28, 2014, 2832-2837.

6. Guerbai, Y., Chibani, Y., **Hadjadji, B.,** 2014, Writer-Independent Handwritten Signature Verification Based on One-Class SVM Classifier. In; International Joint Conference on Neural Networks (IJCNN), Beijing, China, July 6 – 11, 2014, 327-331.

7. **Hadjadji, B.,** Chibani, Y., 2014, Optimized Selection of Training Samples for One-Class Neural Network Classifier, In; International Joint Conference on Neural Networks (IJCNN), Beijing, China, July 6 – 11, 2014, 345-349.

8. Brik, Y., Chibani, Y., **Hadjadji, B.,** Zemouri, E., 2014, Keyword-guided Arabic Word Spotting in Ancient Document Images Using Curvelet Descriptors, In: International Conference on Multimedia Computing and Systems (ICMCS), Marrakech, Morocco, April 14 – 16, 2014.

9. Guerbai, Y., Chibani, Y., **Hadjadji, B.,** 2014, The Effective Use of the One-Class SVM Classifier for Reduced Training Samples and Its Application to Handwritten Signature Verification, In: International Conference on Multimedia Computing and Systems (ICMCS), Marrakech, Morocco, April 14 –16, 2014.

- **National conferences :**

1. **Hadjadji, B**., Chibani, Y., Achiri, L., 2012, The joint Use of Radon Transform and Wigner-Ville Distribution for Skew Angle Estimation and Correction for Printed and Handwritten Images Documents, In: the 3rd International Conference on Image and Signal Processing and their Applications, (ISPA), Mostaganem, Algeria, September, 12-14, 2012.