in Proc. RFIA'2002, Angers, France, vol. 2, pp 499-508, 2002

Un panorama des méthodes de combinaison de classifieurs en reconnaissance de formes
An overview of classifier combination methods in pattern recognition

H. Zouari[1, 2], L. Heutte[1], Y. Lecourtier[1], A. Alimi[2]

[1] Laboratoire Perception, Systèmes, Information (PSI), Université de Rouen, France
[2] Groupe de Recherche sur les Machines Intelligentes (REGIM), Université de Sfax, Tunisie

Laboratoire PSI, UFR des Sciences, Université de Rouen, F-76821 Mont-Saint-Aignan Cedex, France
Hela.Khoufi@univ-rouen.fr

Résumé

Récemment, la combinaison de classifieurs a été proposée comme une voie de recherche permettant de fiabiliser la reconnaissance en utilisant la complémentarité qui peut exister entre les classifieurs. Sur ce point, la littérature abonde de travaux présentant des méthodes de combinaison qui se différencient aussi bien par le type d'informations apportées par chaque classifieur que par leurs capacités d'apprentissage et d'adaptation.

L'objectif de cet article est de dresser un panorama des recherches actuelles sur la combinaison parallèle de classifieurs. Nous rappelons tout d'abord la problématique générale de la combinaison parallèle et proposons une taxonomie des méthodes de combinaison en fonction de critères que nous justifions. Nous détaillons ensuite les méthodes les plus reconnues dans la littérature selon cette taxonomie ainsi que les développements récents dans le domaine. Enfin, nous concluons sur les limites atteintes par ces travaux et en particulier sur la difficulté à généraliser les résultats obtenus en dehors d'un contexte applicatif donné.

Mots Clef

Classifieur, combinaison parallèle de classifieurs, méthode de combinaison, reconnaissance, fusion, sélection.

Abstract

Classifier combination has been recently investigated as a new research field to improve recognition reliability by taking into account the complementarity between classifiers. A large number of combination schemes has been proposed in the literature according to the type of information provided by each classifier as well as their learning and adaptation abilities.

In this communication we present an overview of current research in classifier combination. We first recall what is generally meant by parallel combination and propose a taxonomy of the existing combination methods according to criterions we justify. The most recognized methods are then detailed according to this taxonomy as well as the recent advances in the area. As a conclusion we discuss the limitations of current research and especially the difficulty to generalize the obtained results apart from a specific applicative context.

Keywords

Classifier, parallel combination, classifier combination, recognition, fusion, selection.

1 Introduction

Alors que les premières expériences en combinaison parallèle de classifieurs ne datent que des années 80 [27, 44, 62, 78], cette technique est devenue une voie de plus en plus utilisée [16, 46, 50, 51, 88] pour améliorer la qualité des systèmes de reconnaissance dans plusieurs applications: reconnaissance d'images médicales [18], reconnaissance de chiffres [29, 43, 47, 49, 53], de caractères et de mots manuscrits [40], de visages [1, 11], vérification de signatures [72, 90], reconnaissance de la parole [14, 89], identification de formulaires [19], … Ces systèmes diffèrent par le type de sorties des classifieurs combinés, par la nature des classifieurs utilisés ainsi que par les stratégies de combinaison choisies.

La raison principale pour laquelle la communauté de la reconnaissance de formes porte un véritable intérêt pour la combinaison de classifieurs est sa capacité à prendre en compte un grand nombre de caractéristiques utilisées par des classifieurs différents en exploitant les performances marginales et le comportement de chacun de ces classifieurs. Si, dans un premier temps, les méthodes de classification, basées sur différentes théories et méthodologies, ont été considérées comme autant de solutions possibles à un même problème, leur développement n'a pas permis de mettre en évidence la supériorité incontestable d'une méthode sur une autre pour répondre aux contraintes des applications pratiques. Par conséquent, l'idée d'en utiliser plusieurs simultanément s'est peu à peu imposée. En effet, l'étude de ces techniques a fait apparaître des différences de comportement et donc une complémentarité potentielle qu'il semblait intéressant d'exploiter pour augmenter la robustesse ainsi que la fiabilité de la classification [4, 7,

48, 70, 82]. Ceci s'est traduit par le développement d'un grand nombre de méthodes de combinaison différentes [29, 37, 40, 47, 88].

Cet article est organisé de la manière suivante. Nous commençons par rappeler la problématique de la combinaison parallèle de classifieurs et proposons une taxonomie des méthodes de combinaison (section 2). Nous détaillons dans la section 3 les méthodes de combinaison les plus reconnues dans la littérature selon cette taxonomie ainsi que les développements récents dans le domaine (section 4). Enfin, nous concluons sur les limites atteintes par ces travaux et en particulier sur la difficulté à généraliser les résultats obtenus en dehors d'un contexte applicatif donné.

2 Problématique de la combinaison parallèle

Le problème de la combinaison parallèle de classifieurs peut se poser de la façon suivante : étant donné un ensemble de K classifieurs se prononçant chacun indépendamment sur une même forme à reconnaître, comment élaborer une réponse finale unique à partir des K résultats fournis (figure 1). Ce problème de combinaison parallèle nécessite d'abord de rappeler ce qu'on entend généralement par « classifieur » dans le cadre de la combinaison puis d'examiner les critères à prendre en compte pour catégoriser les différentes méthodes de combinaison présentées dans la littérature.

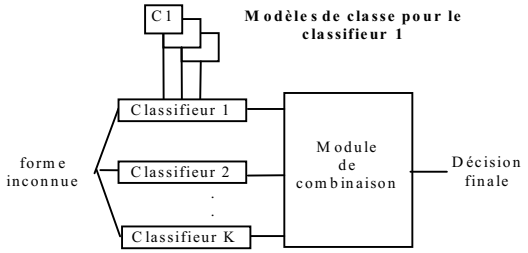

Figure 1: Schéma de combinaison parallèle de classifieurs

2.1 Définition d'un classifieur dans le cadre de la combinaison

Nous appelons *classifieur* tout outil de reconnaissance qui reçoit une forme x en entrée et donne des informations à propos de la classe de cette forme. Cet outil est vu comme une fonction qui, à l'aide des descripteurs de la forme x à reconnaître, décide d'attribuer à x la classe C_i parmi un nombre fini de classes possibles, i=1,...,M.

$$x \rightarrow e(x) = C_i$$

Les réponses fournies par un classifieur peuvent être divisées en trois catégories suivant le niveau d'information apporté par le classifieur :

- type classe : $e_j(x) = C_i$ ($i \in \{1, ..., M\}$), indique que le classifieur j a attribué la classe C_i à x.
- type rang : $e_j(x) = \left\lfloor r_1^j, r_2^j, ..., r_M^j \right\rfloor$ où r_i^j est le rang attribué à la classe i par le classifieur j.
- type mesure : $e_j(x) = \left\lfloor m_1^j, m_2^j, ..., m_M^j \right\rfloor$ où m_i^j est la mesure attribuée à la classe i par le classifieur j.

Chaque type de sortie correspond à un niveau d'information différent fourni par le classifieur. La sortie de type classe est la plus générale mais apporte le moins d'informations. La sortie de type rang reflète l'ordre de préférence des propositions fournies par le classifieur. La sortie de type mesure est la plus riche en information puisqu'elle reflète le niveau de confiance du classifieur dans ses propositions [5]. Toutefois, comme ces mesures ne sont pas toujours comparables (une distance, une probabilité *a posteriori* [28, 81], une valeur de confiance ou une fonction de croyance), une normalisation est souvent nécessaire [42, 63, 88].

2.2 Taxonomie des méthodes de combinaison

Différentes taxonomies des méthodes de combinaison ont été proposées dans la littérature [19, 29, 39, 47, 56, 63, 88].

Dans [56], Kuncheva fait la différence entre fusion et sélection. La fusion consiste à combiner toutes les sorties des classifieurs pour atteindre un consensus [29, 40, 55, 66, 88]. La sélection consiste à choisir dynamiquement les "meilleurs" classifieurs parmi un ensemble de classifieurs possibles pour identifier la forme inconnue [35, 50].

Duin [29] distingue, quant à lui, dans les méthodes de fusion les méthodes de combinaison de classifieurs différents des méthodes de combinaison de classifieurs faibles. En effet, même si ces dernières tiennent compte de toutes les informations produites par les classifieurs, l'intérêt des méthodes de combinaison de classifieurs faibles réside dans la combinaison de classifieurs ayant la même structure mais entraînés sur des données différentes.

Xu [88] distingue les méthodes de combinaison uniquement par le type de sorties des classifieurs (classe, rang, mesure) présentés en entrée de la combinaison. Le type de sorties des classifieurs est un critère qui apparaît dans pratiquement toutes les taxonomies présentées dans la littérature. La prise en compte de ces niveaux d'information différents influencent directement la complexité des méthodes de combinaison développées. Notons que certains auteurs [39, 63] vont même jusqu'à distinguer les méthodes de combinaison de type rang par le type de résultat qu'elles fournissent.

Jain [47] construit une taxonomie suivant trois critères d'égale importance : le type de sorties des classifieurs, l'adaptabilité des méthodes de combinaison (capacité à exploiter l'expertise des classifieurs) et leur capacité d'apprentissage. Ce dernier critère est aussi utilisé par Kuncheva [56] pour séparer les méthodes de fusion. Les méthodes avec apprentissage permettent de chercher et adapter les paramètres à utiliser dans la combinaison suivant la base des exemples disponibles. Les méthodes sans apprentissage se contentent d'utiliser seulement et simplement les sorties des classifieurs sans intégrer

d'autres informations a priori sur les performances de chacun des classifieurs.

La taxonomie des méthodes de combinaison parallèle (figure 2) que nous proposons permet de distinguer au premier niveau les méthodes de fusion des méthodes de sélection. Alors que les méthodes de sélection sont dynamiques ou adaptatives dans la mesure où elles cherchent à combiner le meilleur sous ensemble de classifieurs en fonction des données présentées en entrée de la combinaison, les méthodes de fusion sont statiques puisqu'elles prennent en compte dans un schéma de combinaison fixé toutes les sorties des classifieurs. Ces dernières peuvent être séparées au deuxième niveau suivant la nature des classifieurs combinés : combinaison de classifieurs faibles ou combinaison de classifieurs différents. Si dans le premier cas, les méthodes combinent les résultats de classifieurs identiques (mêmes algorithmes d'apprentissage et de décision) mais entraînés sur des données de distributions différentes, la situation est entièrement différente en ce qui concerne le deuxième type de méthodes dans lesquelles sont combinés des classifieurs qui se différencient aussi bien par leur structure que par les données traitées et les caractéristiques utilisées. Dans les méthodes de combinaison de classifieurs différents, on distingue les méthodes dites « figées » ou sans apprentissage (les sorties des classifieurs sont combinées dans un schéma dont les paramètres sont invariables) des méthodes avec apprentissage qui cherchent à apprendre, sur les données disponibles, les paramètres nécessaires à la combinaison. Enfin, la complexité de ces méthodes peut varier en fonction du niveau d'information associé aux réponses fournies par les classifieurs à combiner (sortie de types classe, rang ou mesure).

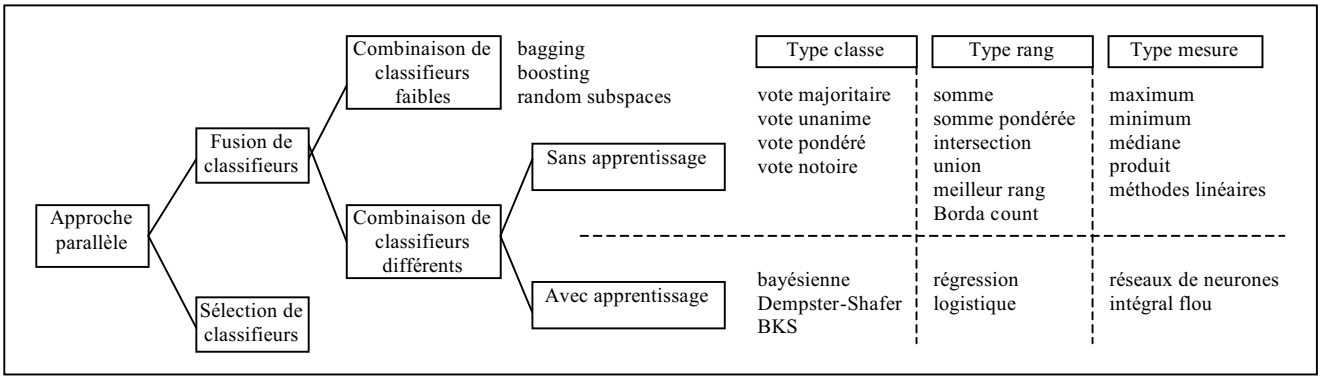

Figure 2 : Taxonomie des méthodes de combinaison parallèle de classifieurs

3 Combinaison de classifieurs différents

3.1 Méthodes sans apprentissage

Nous avons vu que les classifieurs de type classe proposent uniquement la classe d'appartenance de la forme à reconnaître [43, 88]. Dans ce cas, les seules méthodes à appliquer pour combiner ces résultats sans apprentissage sont basées sur le principe de vote. Elles sont très utilisées pour la reconnaissance de l'écrit [17, 31, 40, 51, 57, 64] ainsi que pour la vérification de signatures [72].

Toutes les méthodes de votes peuvent être dérivées de la règle de majorité avec seuil exprimée par :

$$E(x) = \begin{cases} C_i & si \ \sum_i e(i) = \max_{c_i \in \{1,\ldots,M\}} \sum_j e(j) \geq \alpha K \\ rejet & \text{sinon} \end{cases}$$

où K est le nombre de classifieurs à combiner.

Ainsi pour α=1 la classe finale est choisie si tous les classifieurs proposent cette réponse sinon la réponse finale est le rejet. Cette méthode restrictive qui accepte le moins de risque possible est appelée majorité unanime. Pour α=0.5, cela signifie que la classe finale est décidée si plus de la moitié des classifieurs l'ont proposée. Il s'agit alors du vote à la majorité absolue. Pour α=0 il s'agit de la majorité simple où le résultat de la combinaison est la classe la plus proposée parmi les K classifieurs. Néanmoins, il y a rejet si toutes les classes ont le même nombre de votes. Dans ce cas, les risques de conflits sont particulièrement importants. Les résultats de tests [88] montrent que la majorité notoire, qui consiste à choisir une classe comme réponse finale si elle est majoritaire et se distingue de la deuxième classe par une différence supérieure à un certain seuil, est meilleure que la majorité avec seuil surtout quand un minimum de fiabilité est exigé.

La combinaison peut aussi être basée soit sur la majorité sans conflit si tous les classifieurs ont choisi la même classe sauf éventuellement ceux qui ont donné une réponse rejet, soit sur le vote pondéré où la réponse de chaque classifieur est pondérée par un coefficient indiquant son importance dans la combinaison [1, 40].

Dans [1], la forme d'entrée est attribuée à la classe pour laquelle la somme des votes, qui sont pondérés par la fiabilité estimée de chacun des experts, est la plus élevée. Cette méthode n'introduit pas de critère de rejet.

Dieckmann [23] propose un schéma de combinaison basé sur la méthode 2-out-of-3 de vote qui utilise trois classifieurs différents. On pourra notamment citer les travaux de Lam [58] où sont étudiées les propriétés des méthodes de votes et surtout l'apport d'un classifieur supplémentaire dans une combinaison en fonction du nombre de classifieurs présent auparavant (pair ou impair). Cette étude est justifiée si l'on suppose que tous les classifieurs sont indépendants. Srihari [78] présente une analyse théorique de performance de son système qui traite d'un problème de deux classes en utilisant n classifieurs indépendants et le vote majoritaire. Le taux de reconnaissance de chaque classifieur, p, est supposé le même pour tous les classifieurs. Il est montré que le taux de reconnaissance du système de combinaison augmente avec n pour $p > 0.5$, et diminue pour $p < 0.5$.

Nous avons vu que les seules informations utilisées par les méthodes de vote sont les classes indiquées par les classifieurs. Les méthodes de type rang combinent des listes de propositions de classifieurs ordonnées de manière décroissante [39, 73]. Elles utilisent donc plus d'informations sur les classes. Dans cette catégorie de méthodes, on trouve la méthode de somme qui est une généralisation de la méthode de vote. Les rangs attribués par chaque classifieur, peuvent être considérés comme des votes pour toutes les classes. Le vote total pour une classe est donc la somme de tous les votes pour cette classe. Les classes sont ensuite triées selon les rangs (votes) totaux. Le nouvel ordre détermine les rangs finaux. Une extension de cette méthode est la méthode de somme pondérée qui consiste à pondérer les rangs attribués par chaque classifieur par un coefficient indiquant la confiance accordée à ce classifieur. Le choix de ce coefficient de pondération est un problème important. Dans certains travaux [39], le taux de reconnaissance de chaque classifieur est utilisé comme coefficient de pondération. Comme les méthodes précédentes, le Borda count est basé sur le calcul du rang total pour chacune des classes proposées. Il s'agit de faire la somme des nombres de classes placées en dessous d'elle par chacun des classifieurs. Cette méthode est utilisée dans plusieurs travaux de reconnaissance [39, 50, 68, 70] de par son efficacité et sa simplicité. Notons que des variantes de cette méthode se sont révélées plus fiables dans des applications spécifiques de classification [84].

Il existe d'autres méthodes de combinaison de type rang qui n'exigent pas d'apprentissage. Par exemple, la méthode d'intersection, la méthode d'union et la méthode du meilleur rang. La méthode d'intersection consiste à choisir pour chaque classifieur, l'ensemble des meilleurs candidats (les classes de meilleurs rangs) et à réaliser l'intersection entre tous les ensembles obtenus pour la combinaison. Le nombre de classes choisies pour chaque classifieur doit être suffisamment grand pour avoir une probabilité forte de présence de la bonne classe dans l'ensemble de résultat. La combinaison par méthode d'union est réalisée par l'union des meilleurs candidats des classifieurs. Contrairement à l'intersection, le nombre de candidats de chaque classifieur doit être suffisamment petit pour ne pas avoir un ensemble trop important de résultats. La méthode du meilleur rang [39, 73] consiste, tout d'abord, à retenir pour chacune des classes proposées le meilleur rang qui a été affecté par chacun des classifieurs. L'ensemble des classes proposées est ensuite réordonné en fonction de leur meilleur rang. La classe se trouvant finalement en haut de la liste est choisie comme étant la vraie classe. Cette méthode est adaptée pour des problèmes à plusieurs classes et peu de classifieurs. Il apparaît clairement aussi qu'elle peut facilement provoquer des conflits lorsque des classes obtiennent le même meilleur rang. Il est alors difficile de proposer la vraie classe.

Les méthodes de type mesure combinent des mesures qui reflètent le degré de confiance des classifieurs sur l'appartenance de la forme à reconnaître en chacune des classes. Les méthodes de combinaison sans apprentissage dans ce cas sont les méthodes linéaires et les méthodes « fixes ». Dans les méthodes linéaires [39, 42, 82, 87], la règle de décision consiste tout simplement à appliquer aux sorties des classifieurs une combinaison linéaire :

$$E(x) = \sum_{k=1}^{K} \beta_k m_i^k$$

β_k est le coefficient qui détermine l'importance attribuée au $k^{ième}$ classifieur dans la combinaison et m_i^k est sa réponse pour la classe i.

Tumer [82] a montré que la combinaison linéaire de N réseaux de neurones indépendants, non biaisés et ayant des distributions d'erreurs identiques permet de réduire la variance de l'erreur d'un facteur de N. Ce résultat a été établi dans le cas de 2 classes. Dans [83], la réduction d'erreur est obtenue par la combinaison de classifieurs non biaisés ayant des erreurs indépendantes en utilisant une règle moyenne. Alexandre [3] présente une extension de ce travail et montre que la meilleure méthode de combinaison est la moyenne (minimum d'erreur avec 3 bases différentes). Xiao [87] a proposé une méthode linéaire et adaptative pour la reconnaissance de caractères chinois (problème à très grand nombre de classes) et a montré son efficacité par rapport à d'autres méthodes comme le vote pour l'application considérée.

Le principe de base derrière les méthodes fixes [28, 29, 52, 53, 68, 81] est le suivant : les classifieurs sont indépendants et estiment des probabilités *a posteriori* des classes. Plusieurs règles de combinaison peuvent être obtenues pour l'estimation de la probabilité a posteriori en faisant appel à la formule de Bayes. Ainsi, pour reconnaître une forme x en utilisant K classifieurs pour un problème à M classes, les probabilités a posteriori $\{ p_{ij}(x), i = 1, K; j = 1, M \}$ peuvent être calculées par la formule générale :

$$q_j(x) = \frac{q_j'(x)}{\sum_j q_j'(x)}$$

La règle de combinaison dans ce cas est :

$$E(x)=\begin{cases}\max_j\left(q_j(x)\right)\\ rejet \quad \text{sinon}\end{cases}$$

Cette règle de décision finale revient à choisir, parmi l'ensemble des classes les plus probables proposées par chacun des classifieurs, la classe pour laquelle la probabilité a posteriori $q'_j(x)$ est donnée par l'une des règles suivantes :

- maximum : $q'_j(x)=\max_{i=1}^{K}\left(p_{ij}(x)\right)$
- minimum : $q'_j(x)=\min_{i=1}^{K}\left(p_{ij}(x)\right)$
- médiane : $q'_j(x)=med_{i=1}^{K}\left(p_{ij}(x)\right)$
- produit : $q'_j(x)=\prod_{i=1}^{K}\left(p_{ij}(x)\right)$
- somme : $q'_j(x)=\sum_{i=1}^{K}\left(p_{ij}(x)\right)$

Kittler a montré sur une base de chiffres manuscrits que la règle de somme est la plus robuste [53]. Il a confirmé ce résultat par une analyse de l'estimation de l'erreur qui a montré que cette règle est la moins sensible à l'erreur. L'étude menée dans [81] a permis de faire une comparaison entre deux règles de combinaison: médiane et produit. L'auteur a appliqué ces deux techniques en utilisant des classifieurs différents. Il a constaté que si les classifieurs ont une erreur d'estimation faible, les deux règles de combinaison se comportent efficacement quelques soient les primitives utilisées. Lorsque les primitives sont différentes, la combinaison par produit est à préconiser. Il apparaît aussi que dans le cas où l'erreur d'estimation des classifieurs est plus importante, la règle médiane se révèle plus robuste.

3.2 Méthodes avec apprentissage

Contrairement aux méthodes de vote, la théorie de Bayes [9, 26], la théorie de Dempster-Shafer [21, 74] et la méthode d'espace de connaissance du comportement (Behaviour Knowledge Space ou BKS) [43] sont des méthodes de combinaison qui utilisent une phase d'apprentissage pour combiner les résultats de type classe.

L'utilisation de la théorie de Bayes [20, 34, 82], consiste à déterminer la classe C_j pour laquelle la probabilité *a posteriori* $P\left(C_j / e(1)=C_1,...,e(k)=C_k\right)$ est maximum, c'est à dire :

$$E(X)=\begin{cases}C_j \quad si\ \mathrm{P}\left(\mathrm{C_j}/e(1)=C_1,...,e(k)=C_k\right)\\ \qquad =\max_{i=1}^{M}\mathrm{P}\left(\mathrm{C_i}/e(1)=C_1,...,e(k)=C_k\right)\\ rejet\ \text{sin}on\end{cases}$$

Pour estimer la probabilité a posteriori, plusieurs études supposent que les décisions des classifieurs doivent être indépendantes [60, 88]. Sous cette hypothèse et en appliquant la formule d'inversion de Bayes, cette probabilité s'écrit :

$$P\left(C_j / e(1)=C_1,...,e(k)=C_k\right)=P(C_j)\prod_{k=1}^{K}\frac{P(C_j/e(k)=C_k)}{P(C_{ji})}$$

$P(C_j/e(k)=C_k)$ et $P(C_j)$ sont estimées à partir de la matrice de confusion obtenue pour le classifieur k.

Kang [49] estime la probabilité a posteriori en utilisant une approche de dépendance de produit. Il cherche l'approximation de produit « optimale » pour la distribution de probabilité d'ordre élevé sans proposition d'hypothèses d'indépendance. La méthode proposée par Moobed [63] est basée sur une approche bayésienne. Cette méthode est justifiée théoriquement et présente plusieurs avantages. En effet, elle utilise toutes les informations disponibles, y compris le comportement des classifieurs sur différentes classes. Elle prend en considération tout type de dépendance entre les classifieurs. L'inconvénient majeur de cette méthode est la sensibilité aux nombre d'exemples disponibles.

Le concept central de la théorie de Dempster et Shafer [21, 74] est celui de la structure de croyance $m : 2^{\Omega} \rightarrow [0, 1]$ qui vérifie les propriétés suivantes :

$$m(\emptyset)=0 \quad \text{et} \quad \sum_{A\subseteq\Omega}m(A)=1$$

où $\Omega=\{\omega_1,...,\omega_M\}$, $2^{\Omega}=\{A=\{\omega_{i1},...,\omega_{in}\}/A\subseteq\Omega\}$.

$m(\mathrm{A})$ correspond au degré d'évidence attribuée strictement à l'ensemble A, mais ne peut être associée à aucun sous ensemble particulier de A. Les éléments focaux de m sont les sous-ensembles A de Ω pour lesquels la masse de croyance $m(\mathrm{A})$ est non nulle.

Dans DS, la méthode de combinaison des évidences $m_1,...m_k$ issues de plusieurs classifieurs supposés indépendants est: $m=m_1\oplus m_2\oplus...\oplus m_k$

Etant donné un élément focal A, $m(A)$ est obtenue par la somme des produits des mesures d'évidence issues des éléments $A_1,...,A_k$ dont l'intersection correspond au sous-ensemble A :

$$m(A)=\frac{1}{C}\sum_{A1\cap...\cap Ak=A}\prod_{i=1}^{k}m_i(A_i)$$

La quantité $C=1-\sum_{A_1\cap..\cap A_k=\phi}\prod_{i=1}^{k}m_i(A_i)$ est un facteur de normalisation qui mesure le conflit entre les différentes structures de masse.

L'utilisation de la règle de combinaison [22, 36, 61, 62, 71] a longtemps été limitée à des sources indépendantes. Mais Quinion [69] a prouvé sa validité dans le cas de sources dépendantes. Une application de la combinaison d'évidence est fournie dans [91] où l'affectation des masses de croyance est une fonction inverse de la distance au plus proche voisin. Basée sur la même règle de combinaison, Xu [88] a utilisé les performances des classifieurs pour réaliser l'affectation des masses de croyance.

Si les deux méthodes de combinaison exposées précédemment prennent en compte le comportement a priori des classifieurs, elles sont néanmoins basées sur

l'hypothèse forte d'indépendance entre classifieurs. Cette hypothèse n'est pas nécessaire dans la méthode BKS [43, 56, 68]. En effet, cette méthode permet de prendre en compte les dépendances en les intégrant directement dans la règle de décision. Elle utilise un espace de connaissances qui contient toutes les décisions des classifieurs ce qui permet de connaître leur comportement. A partir de cette connaissance, il est alors possible d'obtenir la décision finale.
BKS est un espace à N dimensions où chaque dimension correspond à la décision d'un classifieur. Un exemple de BKS à 2 dimensions est présenté dans le tableau 1.

e2 \ e1	1	...	j	...	11
1	(1,1)		(1,j)		(1,11)
.	...		...		...
i	...		(i,j)		...
.	...		...		...
11	(11,1)		(11,j)		(11,11)

Tableau 1 : BKS de 2 classifieurs

L'intersection des décisions des classifieurs individuels occupe une unité de l'espace BKS(e(1), ..., e(K)) appelée *unité focale.* Chacune comporte 3 types d'informations :

- le nombre total d'échantillons qui appartiennent à la classe C_i : $n_{e(1)...e(K)}(C_i)$
- le nombre total d'échantillons présents dans l'unité : $T_{e(1)...e(K)} = \sum_{m=1}^{M+1} n_{e(1)...e(K)}(C_i)$ (1)
- la classe la plus représentée dans l'unité : $R_{e(1)..e(K)} = \left\{ C_j \setminus n_{e(1)..e(K)}(C_j) = \max_{i=1}^{M+1} (n_{e(1)..e(K)}(C_i)) \right\}$ (2)

La méthode BKS est constituée de deux étapes : une étape d'apprentissage et une étape de décision. Au cours de l'apprentissage, BKS est construit à partir des résultats fournis par les K classifieurs : les valeurs $n_{e(1)...e(K)}(C_i)$ sont déterminées après un passage complet de la base d'apprentissage, les valeurs de $T_{e(1)...e(K)}$ et $R_{e(1)..e(K)}$ de chaque unité de BKS(e(1), ..., e(K)) sont ensuite calculées en utilisant (1) et (2) respectivement. L'étape de décision repose sur l'utilisation des réponses des K classifieurs à la forme présentée pour choisir l'unité focale correspondante. La règle de décision suivante est ensuite utilisée pour donner le résultat final à propos de la classe x :

$$E(x)= \begin{cases} R_{e(1)...e(K)}, si\ T_{e(1)...e(K)} > 0 \\ \qquad et \quad \dfrac{n_{e(1)...e(K)}(R_{e(1)...e(K)})}{T_{e(1)...e(K)}} > \lambda \\ M+1, \text{ sinon} \end{cases}$$

où λ est un seuil qui contrôle la fiabilité de la décision finale ($0< \lambda <1$)
L'étude réalisée dans [43] montre l'efficacité de cette méthode. Pour les applications pratiques, l'utilisation de BKS dépend de la base d'apprentissage qui doit être représentative et riche.

Comme les méthodes précédentes, la méthode de régression logistique [73, 85] utilise des informations sur le comportement des classifieurs mais combinent des sorties de type rang. Elle est introduite dans [39] comme une méthode de généralisation de *Borda Count* en utilisant les poids comme information reflétant l'importance relative des classifieurs. Cette méthode utilise la probabilité $\pi(D_i)$ indiquant que la classe finale de x est la classe C_i avec $0 < \pi(D_i) < 1$ et D_i est le vecteur des réponses de classifieurs pour la classe C_i. Pour l'estimation de cette probabilité, on utilise la régression logistique qui a la forme suivante :

$$\pi(D_i) = \frac{\exp\ (\beta_0 + \beta_1 r_1 + \beta_2 r_2 + ... + \beta_K r_K)}{1 + \exp\ (\beta_0 + \beta_1 r_1 + \beta_2 r_2 + ... + \beta_K r_K)} \quad (3)$$

Les paramètres de régression β_i peuvent être estimés par les méthodes des moindres carrés ou du maximum de vraisemblance [2].
Une fois, les paramètres β_i calculés, la combinaison des K vecteurs de rang fournis par les classifieurs consiste à calculer les probabilités $\pi(D_i)$ en utilisant l'équation (3). Les classes finales de la combinaison seront les classes triées selon les valeurs de $\pi(D_i)$. Finalement, la classe ayant la probabilité la plus élevée est sélectionnée.

D'autres méthodes qui ont bénéficié de l'essor des travaux sur la classification dans plusieurs domaines sont utilisées pour combiner les classifieurs de type mesure : l'intégral flou et les réseaux de neurones. Le principe de l'intégral flou est de calculer des fonctions floues pour chacune des classes et retenir la classe ayant l'intégral flou le plus élevé [8, 12, 15, 75, 79]. L'utilisation d'un réseau de neurones en tant que méthode de combinaison est particulièrement séduisante [6, 11, 12, 13, 38, 60]. Grâce à sa capacité d'apprentissage, le réseau doit pouvoir gérer au mieux les conflits qui apparaissent entre les classifieurs. Par exemple, pour corriger les erreurs inhérentes de 2 classifieurs (kppv), Prevost [67] utilise la fusion neuronale lui permettant ainsi d'améliorer notablement les performances d'un système de reconnaissance de caractères. On peut citer aussi le travail de Huang [42] dans lequel le réseau de neurones est utilisé pour combiner 3 classifieurs après avoir transformé leurs sorties en mesure de même type.

4 Développements récents en combinaison

Les travaux récents dans le domaine de la combinaison parallèle traite de l'optimisation des systèmes de décision, soit en cherchant à combiner des classifieurs de faibles performances, soit en sélectionnant dynamiquement un sous ensemble minimal de classifieurs parmi un ensemble de classifieurs à combiner.

4.1 Combinaison de classifieurs faibles

Pour améliorer le résultat de classifieurs faibles, certains chercheurs [33, 77] utilisent une des trois techniques de

combinaison : Bagging [76], Boosting [32] et/ou random subspaces [41].

Bagging [76] est basée sur des concepts de **B**ootstrapping [30] et d'**ag**gre**ging**. Le bootstrapping est conçu pour générer au hasard et avec remise K copies indépendantes de N objets $X^b = (X_1^b,..., X_n^b)$ appelées *bootstrap* à partir de l'ensemble initial des échantillons d'apprentissage de n objets $X = (X_1,..., X_n)$ et pour construire un classifieur avec chacune de ces copies. On obtient par la suite K classifieurs de différentes performances entraînés sur chacune de ces copies. L'agrégation consiste à combiner ces classifieurs en utilisant le vote majoritaire comme règle de combinaison.
L'étude simulée de Skurichina [76] a montré que, généralement, le bagging est une technique de combinaison qui permet d'améliorer la performance des classifieurs linéaires instables. Nous trouvons la même conclusion dans [10] où l'auteur note que la réduction de l'erreur de classification ne peut être obtenue que dans le cadre d'une combinaison de classifieurs instables. Pour des classifieurs stables, elle détériore la performance de la combinaison.
Différentes modifications du bagging ont été étudiées [80]. En particulier, la technique dite du 'nice bagging' qui consiste à combiner les meilleurs classifieurs (c'est à dire les classifieurs dont les erreurs obtenues sur les versions de bootstrap d'apprentissage sont les plus faibles) est préférable au bagging, dans la mesure où elle permet de construire le classifieur avec le minimum d'erreur de classification [76].

Contrairement au bagging où les ensembles d'apprentissage de bootstrap et les classifieurs sont construits de manière indépendante, dans boosting [32], les échantillons d'apprentissage sont construits incrémentalement par le même classifieur et de manière séquentielle. Initialement, tous les échantillons d'apprentissage ont des poids égaux et le classifieur est construit sur cette base. Ensuite pour chaque étape, les échantillons sont pondérés de façon à ce que les objets mal classés (qui se trouvent généralement aux frontières des classes) aient des poids élevés et le classifieur est lancé sur le nouvel ensemble d'apprentissage ainsi pondéré. De cette manière, on obtient finalement un ensemble de classifieurs qui sont combinés par un vote majoritaire ou pondéré pour avoir la décision finale. D'autres versions du boosting sont étudiées par certains chercheurs [25, 48] pour combiner des classifieurs faibles (réseaux de neurones). Drucker [25] a montré que certaines versions du boosting sont préférables dans le cas où la taille des données à apprendre est grande. Cependant, par son coût de calcul plus faible, l'algorithme de boosting original reste le meilleur. L'algorithme proposé par Ji [48] est testé sur des bases artificielles et réelles en fonction de certains paramètres (le nombre de classifieurs, le facteur de faiblesse permettant de sélectionner les classifieurs faibles à combiner et le seuil de partition de la base d'apprentissage). L'analyse théorique présenté par Ji a permis de fournir des solutions pour le problème de choix du facteur de faiblesse et le seuil de partition de la base d'apprentissage. Cependant le problème du choix du nombre de classifieurs reste un problème ouvert.
Contrairement à l'étude de Friedman [33] où il est difficile d'utiliser d'autres règles de combinaison à part le vote simple ou pondéré, celle de Skurichina [77] fait appel à d'autres méthodes de combinaison, à savoir l'analyse de discrimination linéaire (LDA), la règle moyenne et la règle du produit. Les résultats de cette étude ont montré que le boosting est meilleur que le bagging pour toutes les méthodes de combinaison utilisées et sur des bases d'apprentissage de grande taille.

Une autre technique de combinaison permettant d'améliorer la performance des classifieurs faibles est la méthode « random subspaces » (RSM) [41, 65]. En effet, elle consiste à modifier la base d'apprentissage comme dans les deux autres techniques (bagging et boosting). Cependant, cette modification est réalisée sur l'espace de caractéristiques. Soit $X_i = (x_{i1}, x_{i2},..., x_{ip})$ un élément de dimension p de la base d'apprentissage $X = (X_1,..., X_n)$. On sélectionne au hasard $p^* < p$ caractéristiques de *X*. L'ensemble d'apprentissage modifié $\tilde{X}^b = (\tilde{X}_1^b, \tilde{X}_2^b,..., \tilde{X}_n^b)$ est constitué d'objets $\tilde{X}_i^b = (x_{i1}^b, x_{i2}^b,..., x_{ip^*}^b)$ (i=1,...,n) de dimension p^* où les x_{ij}^b (j=1,...,p*) sont les p* composantes du vecteur d'apprentissage X_i sélectionnées au hasard. Les classifieurs sont ensuite construits dans les sous-espaces $\tilde{X}^b$ et combinés par la règle de vote majoritaire.
Skurichina [77] a montré que la méthode RS permet de maintenir une erreur faible en apprentissage et d'améliorer l'erreur de généralisation pour des classifieurs linéaires. Elle a constaté que la méthode RS peut être plus performante que le bagging et le boosting si le nombre de caractéristiques est élevé.

4.2 Sélection de classifieurs

Ce n'est que récemment que les travaux sur la combinaison ont montré l'importance de traiter le problème de sélection des classifieurs [24, 35, 40, 50]. Cette approche considère chaque classifieur comme un expert dans son espace local de caractéristiques. La décision finale peut être prise soit uniquement par le meilleur classifieur [35], soit par plusieurs classifieurs en même temps [46, 59]. Woods [86] montre qu'il existe certaines stratégies de sélection de classifieurs mixtes qui permettent d'obtenir de meilleures performances que si on utilise tous les classifieurs disponibles. Il considère donc une simple recherche séquentielle pour la sélection de classifieurs et propose une procédure qui détermine si ces

classifieurs sont redondants ou « nuisibles ». La méthode de sélection proposée dans [35] consiste à regrouper dans un vecteur tous les réponses des classifieurs pour une forme de test, rechercher l'ensemble des vecteurs de la base d'apprentissage qui ressemble à ce vecteur et calculer la précision de chaque classifieur à partir de cet ensemble. Le classifieur ayant la précision la plus élevée est choisi pour identifier la forme de test. Impedovo [45] propose une nouvelle méthodologie qui cherche le meilleur ensemble de classifieurs pour la combinaison. Cette méthodologie utilise l'indice de similarité pour estimer la corrélation stochastique de l'ensemble de classifieurs (mesure de l'accord entre les sorties des classifieurs). Cet indice est aussi utilisé dans [24, 50, 59] pour la sélection de classifieurs ainsi que pour l'évaluation des méthodes de combinaison.

5 Conclusion

A notre connaissance, tous les résultats présentés dans la littérature montrent clairement que la combinaison parallèle de classifieurs est une voie de recherche prometteuse. Elle permet d'améliorer les performances du système de reconnaissance et de s'adapter à une grande variété de situations par l'exploitation de connaissances à priori sur les comportements des classifieurs.

Les travaux effectués sur la combinaison montrent aussi la variété des méthodes de combinaison qui diffèrent par leur capacité d'apprentissage, le type de sortie des classifieurs et la manière de les choisir (sélection ou non). Toutefois, ces travaux qui peuvent être optimisés pour des problèmes donnés [16, 66] sont difficilement généralisables en dehors d'un domaine applicatif donné. En effet, le développement d'un système de combinaison ne suit a priori aucune règle précise, et dépend étroitement de l'application que l'on veut traiter, de la façon dont on veut la traiter et des outils disponibles (bases de données, classifieurs et règles de combinaison). Ces contraintes expliquent bien les limites des systèmes développés du point de vue performance pour traiter d'autres applications plus complexes.

Par opposition, le nombre d'études théoriques [54, 57] qui traitent de la combinaison d'un point de vue général, sans s'attacher à une application particulière, ni aux types de sorties des classifieurs, reste très faible. Bien qu'ils abordent les vrais problèmes de combinaison, ils utilisent des hypothèses rarement vérifiées.

Il est donc clair que la diversité des travaux sur la combinaison n'a pas permis d'avoir une approche générale qui traite le problème de la combinaison de manière satisfaisante dans tous les cas de figure. La mise en oeuvre d'une telle approche n'ira pas sans poser certaines questions : Comment modéliser la combinaison des informations produites par les classifieurs et exploiter leur complémentarité pour un problème donné ? En fonction de quels critères et comment l'évaluer ? Comment évaluer la robustesse d'une méthode donnée sur des applications différentes ? Ces questions restent des problèmes ouverts de reconnaissance et constituent un défi à relever dans les années futures.

6 Bibliographie

[1] B. Achermann & H. Bunke, "Combination of classifiers on the decision level for face recognition", Rapport technique, university of Bern, 1996.

[2] A. Agesti, "Categorical data analysis", John Wiley & Sons, 1990.

[3] L.A. Alexandre, A.C. Campilho & M. Kamel, "Combining independant and unbiased classifiers using weighted average", ICDAR'00, 2000.

[4] K. Anisimovich, V. Rybkin, A. Shamis & V. Tereshchenko, "Using combination of structural, feature and raster classifiers for recognition of handprinted characters", ICDAR'97.

[5] A.S. Atukorale & P.N. Suganthan, "Combining classifiers based on confidence values", Proceedings of ICDAR'99, 5th International Conference on Document Analysis and Recognition, pp. 37-40, India, September 1999.

[6] R. Bajaj & S. Chaudhury, "Signature verification using multiple neural classifiers ", Pattern Recognition, Vol 30, no 1, pp. 1-7, 1997.

[7] R. Battati & A.M. Colla, "Democracy in neural nets : voting schemes for classification", Neural Networks, Vol 7, n 4, pp. 691-707, 1994.

[8] S. Beiraghi, M. Ahmadi, M. Shridhar & M.S. Ahmed, "Application of fuzzy integrals in fusion of classifiers for low error rate handwritten numerals recognition", ICPR'00, 2000.

[9] J.O. Berger, "Statistical decision theory and bayesian analysis", Springer-Verlag, 1985.

[10] L. Breiman, "Bagging predictors", Machine Learning, vol 24, no 2, pp. 123-140, 1996.

[11] R. Brunelli & D. Falavigna, "Person identification using multiple cues", IEEE Trans. Pattern Anal. Mach. Intell., vol 17, no 10, pp. 955-966, 1995.

[12] J. Cao, M. Ahmadi & M. Shridhar, "Fusion of classifiers with fuzzy integrals", ICDAR'95, 1995.

[13] Z. Chi, M. Suters & H. Yan, "Handritten digit recognition using combined ID3-derived fuzzy rules and Markov chains", Pattern Recognition, vol 29, no 11, pp. 1821-1833, 1996.

[14] C.C. Chibelushi, J.S. Mason & F. Deravi, "Integration of acoustic and visual speech for speaker recognition", EUROSPEECH'93, pp. 157-160, 1993.

[15] S-B. Cho & J.H. Kim, "Combining multiple neural network by fuzzy integral for robust classification", IEEE Transaction on Systems, Man, And Cybernetics, Vol 25, No 2, February 1995.

[16] S-B. Cho, "Pattern recognition with neural networks combined by genetic algorithm", Fuzzy Sets and Systems 103, pp. 339-347, 1999.

[17] K.H. Chou, L.T. Tu & I.S. Shyu, "Performance analysis of a multiple classifiers system for recognition of totally unconstrained handwritten numerals", 4th Internationanl Workshop on Frontiers of Handwriting Recognition, pp. 480-487, Taiwan, 1994.

[18] Y.Y. Chou & L.G. Shapiro, "A hierarchical multiple classifier learning algorithm", ICPR'2000.

[19] E. Clavier, E. Trupin, M. Laurent, S. Diana & J. Labiche, "Classifiers combination for forms sorting", ICPR'00, 2000.

[20] L.P. Cordella, P. Foggia, C. Sansone, F. Tortorella & M. Vento, "Optimizing the error/reject trade-off for a multi-

expert system using the bayesian combining rule", advances in pattern recognition 1998.

[21] A.P. Dempster, "Upper and lower probabilities induced by a multivalued mapping", Annals of Mathematical Statistics, AMS, vol 38, pp. 325-339, 1967.

[22] T. Denoeux, "Analysis of evidence theoretic decision rules for pattern classification", Pattern Recognition, Vol 30, No 7, pp. 1095-1107, 1997.

[23] U. Dieckmann, P. Plankensteiner & T. Wagner, "Sesam : A biometric person identification system using sensor fusion", Pattern Recognition letters, Vol 18, No 9, pp. 827-833, 1997.

[24] G. Dimauro, S. Impedovo, G. Pirlo & S. Rizzo, "Multiple experts : a new methodology for the evaluation of the combination processes", the 5th Inter. Work. On Frontiers in Hand. Reco., pp. 131-136, 1995.

[25] H. Drucker, C. Cortes, L. Jackel, Y. Lecun & V. Vapnik, "Boosting and other ensemble methodes", Neural Computation, Vol 6, no 6, pp. 1289-1301, 1994.

[26] R.O. Duda & P.E. Hart, "Pattern classification and scene analysis", John Wiley Sons, 1973.

[27] B. Duerr, W. Haettich, H. Tropf & G. Winkler, "A combination of statistical and syntactical Pattern recognition applied to classification of unconstrained handwritten numerals", Pattern Recognition, vol 12, pp. 189-199, 1980.

[28] R.P.W. Duin & D.M.J. Tax, "Classifier conditional posterior probabilities", Advances in Pattern Recognition, Vol 1451, Berlin, pp. 611-619, 1998.

[29] R.P.W. Duin & D.M.J. Tax, "Experiments with classifier combinig rules", Multiple classifier systems (Proc. First Int. Workshop, MCS 2000), vol 1857, Springer, Berlin, pp. 16-29, 2000.

[30] B. Efron & R. Tibshirani, "An introduction to the bootstrap", Chaman and Hall, 1993.

[31] J. Franke & E. Mandler, "A comparison of two approaches for combining the votes of cooperating classifiers", 11th IAPR International Conference on Pattern Recognition, pp. 611-614, Den Haag, Netherlands, 1992.

[32] Y. Freund & R.E. Schapire, "Experiments with a new boosting algorithm", Machine Learning : Proceedings of the 13th International Conference, July 1996.

[33] J.H. Friedman, "Regularized discriminant analysis", J.Am. Statistical Assoc, vol84, pp. 165-175, 1989.

[34] P. Foggia, C. Sansone, F. Tortorelle & M. Vento, " Multi-classification : rejet criteria for the bayesian combiner ", Pattern Recognition, vol 32, no 8, pp. 1435-1447, 1999.

[35] G. Giacinto, F. Roli & G. Fumera, "Selection of classifiers based on multiple classifier behaviour", Advances in Pattern Recognition, septembre 2000.

[36] V. Gunes M. Ménard & P. Loonis, "A multiple classifier system using ambiguity rejection for clustering-classification cooperation", Internat. Journal of Uncertaity, Fuzziness and Knowledge-Based Systems, vol 8, no 6, 2000.

[37] L.K. Hansen & P. Salamon, "Neural network ensembles", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol 12, No 10, pp. 993-1001, October, 1990.

[38] H-W. Hao, X-H. Xiao & R-W. Dai, "Handwritten Chinese character recognition by metasynthetic approach", Pattern recognition, vol 30, no 8, pp. 1321-1328, 1997.

[39] T.K. Ho, "A theory of multiple classifier systems and its application to visual word recognition", PhD thesis of the university of New-York at Bufallo, 1992.

[40] T.K. Ho, J.J. Hull & S.N. Srihari "Decision combination in multiple classifier systems", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol 16, no 1, pp. 66-75, january 1994.

[41] T.K. Ho, "The random subspace method for constructing decision forests", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol 20, No 8, August 1998.

[42] Y.S. Huang, K. Liu & C.Y. Suen "The combination of multiple classifiers by neural network approach", International Journal of Pattern Recognition and Artificial Intelligence, Vol 9,No 3, pp. 579-597, 1994.

[43] Y.S. Huang & C.Y. Suen "A method of combining multipe experts for the recognition of unconstrained handwritten numerals", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol 17, no 1, pp. 90-94, 1995.

[44] J.J. Hull, "A comutational theory of visual word recognition", Doctoral Dissertation, Technical Report 88-07, Department of computer science, SUNY at Buffalo, February 1988.

[45] S. Impedovo & A. Salzo, "Evaluation of combination methods", Proceedings of ICDAR'99, 5th International Conference on Document Analysis and Recognition, pp. 394-397, Bangalore, India, September 1999.

[46] R.A. Jacobs, M. Jordan, S. Nowlan & G. Hinton, "Adaptive mixtures of local experts", Neural computation, Vol 3, no 1, pp. 79-87, 1991.

[47] A.K. Jain, R.P.W. Duin & J. Mao, "Statistical pattern recognition :A Review", IEEE Transaction on Pattern Analysis and Machine Intelligence, vol 22, no 1, january 2000.

[48] C. Ji & S. Ma, "Combination of week classifiers", IEEE Trans. Neural Networks, vol 8, no 1, pp. 32-42, 1997.

[49] H.J. Kang & J.H. Kim, "A probabilistic framework for combining multiple classifiers at abstract level", Proc. Of ICDAR'97, 4th International conference on Document Analysis and recognition, Germany, vol 2, pp. 870-874, Aug. 1997.

[50] J. Kim, K. Seo & K. Chung, "A systematic approach to classifier selection on combining multiple classifiers for handwritten digit recognition", ICDAR 1997.

[51] F. Kimura & M. Shridhar "Handwritten numerical recognition based on multiple algorithmes", Pattern Recognition, vol 24, no 10, pp. 969-983, 1991.

[52] J. Kittler, A. Hojjatoleslami & T. Windeatt, "Strategies for combining classifiers employing shared and distinct pattern representations", Pattern Recognition Letters, Vol 18, pp. 1373-1377, 1997.

[53] J. Kittler, M. Hatef, R.P.W. Duin & J. Matas, "On combining classifiers", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol 20, No 3, March 1998.

[54] R.M. Kleinberg, "Stochastic discrimination", Annals of Math, and Artificial Intelligence, vol 1, pp. 207-239, 1990.

[55] L.I. Kuncheva, C.J. Whitaker, C.A. Shipp & R.P.W. Duin, "Is independence good for combining classifiers ? ", ICPRI 5, Proc. 15th Inter. Conf. On Pattern Recognition, vol 2, pp. 168-171, 2000.

[56] L.I. Kuncheva, C.J. Bezdek & R.P.W. Duin, "Decision templates for multiple classifier fusion : on experimental comparison", Pattern Recognition, 34, pp. 299-314, 2001.

[57] L. Lam & C.Y. Suen, "A theoretical analysis of the application of majority voting to pattern recognition", International Conference on Pattern Recognition, pp. 418-420, Jerusalem, 1994.

[58] L. Lam & C.Y. Suen, "Application of majority voting to pattern recognition : an analysis of its behavior and performance", IEEE Trans. On Syst., Man, and Cybernetics, Vol 27, No 5, Sepember 1997.

[59] V. Di Lecce, G. Dimauro, A. Guerriero, S. Impedovo, G. Pirlo & A. Salzo, "Classifier combination : the role of a-priori knowledge", Proceedings of the 7th International Workshop on Frontiers in Handwriting Recognition, pp. 143-152, September 2000.

[60] D.S. Lee & S.N. Srihari, "A theory of classifier combination : the neural network approach", Proc. Of ICDAR'95, 3rd International Conference on Document Analysis and Recognition, Montréal, Canada, pp. 42-45, Aug. 1995.

[61] P. Loonis, M. Ménard & J.P. Bonnefoy, "Fusion d'information multi-sources : Etude comparative entre une approche connexionniste dirigée et règle orthogonale de Dempster-Shafer", Proc Of RFIA'96, Rennes, France, pp. 606-614, 1996.

[62] E. Mandler & J. Schuermann, "Combining the classification results of independent classifiers based on the dempster-shafer theory of evidences", International Journal of pattern Recognition and Artificial Intelligence, pp. 381-393, North-Holland, 1988.

[63] B. Moobed, "Combinaison de classifieurs : une nouvelle approche", Thèse de Doctorat, Laboratoire LIX, Ecole Polytechnique, Palaiseau, 1996.

[64] C. Nadal, R. Legault & C.Y. Suen, "Complementary algorithms for the recognition of totally unconstrained handwritten numerals", 10th International Conference on Pattern Recognition, Atlantic City, pp. 443-449, 1990.

[65] E. Pekalska, M. Skurichina & R.P.W. Duin, "Combining Fisher linear discriminants for dissimilarity representations", Multiple classifier systems, MCS, vol 1857, pp. 117-126, 2000.

[66] T. Pham & H. Yan, "Fusion of handwritten numeral classifiers based on fuzzy and genetic algorithms", In : Proceedings of the 1997 Annual Meeting of the North American Fuzzy Information Processing Society, NA-FIPS'97, New York, pp. 257-262, 1997.

[67] L. Prevost & M. Milgram, "Coopérations pour la reconnaissance de caractères dynamiques isolés", 11ième Congrès RFIA'98, vol 3, pp. 233-240, 1998.

[68] S. Procter & J. ILLingworth, "Combining HMM classifiers in a handwritten text recognition system", IEEE International Conference on Proceesing, Chicago, 1998.

[69] P. Quinion & T. Matsuyama, "Random closed sets : a unified approach to the representation of imprecision and uncertaity", In European Conf. On Symb. And Quant. Appro. To Uncer., pp. 282-286, France, 1991.

[70] A.F.R. Rahman & M.C. Fairhurst, "Multiple expert classification : a new methodology for parallel decision fusion", International Journal on Document Analysis and Recognition, IJDAR, Vol 3, pp. 40-55, 2000.

[71] G. Rogova, "Combining the results of several neural network classifiers", Neural Networks, vol 7, no5, pp. 777-781, 1994.

[72] M. Sabourin & G. Genest, "Coopération de classificateurs pour la vérification automatique des signatures", 3ème Colloque National sur l'Ecrit et le document, pp89-98, Rouen, 1994.

[73] A. Saranli & M. Demirekler, "A statistical unified framework for rank-based multiple classifier decision combination", Pattern Recognition, Vol 34, pp. 865-884, 2001.

[74] S. Shafer, "A mathematical theory of evidence", Princeton University Press, 1976.

[75] M. Signahl & A. Verikas, "Fuzzy combination schemes for modular neural networks", the 10th Scandinavian Conference On Image Analysis, Finland, 1997.

[76] M. Skurichina & R.P.W. Duin, "Bagging for linear classifiers", Pattern Recognition, Vol 31, No 7, pp. 909-930, 1998.

[77] M. Skurichina & R.P.W. Duin, "Boosting in linear discriminant analysis", Multiple classifiers systems, MCS, Vol 1857, pp. 190-199, 2000.

[78] S.N. Srihari, "Reliability analysis of majority vote systems", Information Sciences, Vol 26, pp. 243-256, 1982.

[79] M. Sugeno, "Fuzzy measures and fuzzy integrals : a survey", Fyzzy Automata and Decision Processes, Amsterdam : North Holland, pp. 89-102, 1977.

[80] M. Taniguchi & V. Tresp, "Averaging regularized estimators", Neural Computation, vol 9, pp. 1163-1178, 1997.

[81] C.M.J. Tax, R.P.W. Duin & M.V. Breukelen, "Comparison between product and mean classifier combination rules", Proc. Int. Workshop Statistical techniques in Pattern Recognition, pp. 165-170, 1997.

[82] K. Tumer & J.D. Ghosh, "Estimating the bayes error rate through classifier combining", ICPR'96, pp. 695-699, 1996.

[83] K. Tumer & J.D. Ghosh, "Analysis of decision boundaries in linearly combined neural classifiers", Pattern recognition vol 29, no 2, pp. 341-348, 1996.

[84] M. Van Erp & L. Schomaker, "Variants of Borda Count method for combining ranked classifier hypotheses", Proceedings of the 7th International Workshop on Frontiers in handwriting Recognition, september 11-13, 2000.

[85] P. Verlinde, G. Chollet & M. Acheroy, "Multi-modal identity verification using expert fusion", Information Fusion, pp. 17-33, 2000.

[86] K. Woods, W.P Kedelmeyer & K. Bowyer, "Combination of multiple classifiers using local accuracy estimates", IEEE PAMI, vol 19, no 4, pp. 90-94, 1997.

[87] B.H. Xiao, C.H. Wang & R.W. Dai, "Adaptive combination of classifiers and its application to handwritten Chinese charcter recognition", ICDAR'2000.

[88] L. Xu, A. Krzyzak & C.Y. Suen "Methods of combining multiple classifiers and their applications to handwriting recognition", IEEE Transaction on Systems, Man, And Cybernetics, vol 22, no 3, pp. 418-435, may/june 1992.

[89] Yu, X. Jiang & H. Bunke, "Combining acoustic and visual classifiers for the recognition of spoken sentences", ICPR'00, 2000.

[90] E.N. Zois & V. Anastassopoulos, "Fusion of correlated decisions for writer verification", Pattern Recognition, vol 32, pp. 1821-1823, 1999.

[91] L.M. Zouhal, "Contribution à l'application de la théorie des fonctions de croyance en reconnaissance des formes", Thèse de doctorat de l'université de Compiègne, 1997.