

# Une méthode de sélection de caractéristiques fondée sur l'intégrale de Choquet et l'analyse des typicalités

C. MAZAUD<sup>1</sup>J. RENDEK<sup>2</sup>V. BOMBARDIER<sup>1</sup>L. WENDLING<sup>2</sup>

<sup>1</sup>CRAN, Centre de Recherche en Automatique de Nancy, CNRS UMR 7039, Université Henri Poincaré – Campus scientifique, BP 239, 54506 Vandoeuvre-lès-Nancy Cedex

<sup>2</sup>LORIA, Laboratoire Lorrain de Recherche en Informatique et ses Applications, UMR 7053, Université Henri Poincaré – Campus scientifique, BP 239, 54506 Vandoeuvre-lès-Nancy Cedex

[cyril.mazaud@cran.uhp-nancy.fr](mailto:cyril.mazaud@cran.uhp-nancy.fr)

[jan.rendek@loria.fr](mailto:jan.rendek@loria.fr)

[vincent.bombardier@cran.uhp-nancy.fr](mailto:vincent.bombardier@cran.uhp-nancy.fr)

[laurent.wendling@loria.fr](mailto:laurent.wendling@loria.fr)

**Résumé** – Cet article expose une méthode itérative de sélection de caractéristiques fondée sur l'intégrale de Choquet et l'analyse des typicalités. La méthode proposée est appliquée à la reconnaissance de défauts sur des planches de bois, domaine dans lequel peu de données sont disponibles en apprentissage. Le choix d'un sous-ensemble de caractéristiques parmi toutes celles que l'on peut extraire d'une image n'est pas aisé. Il apparaît, à l'usage, que même un expert du domaine choisit souvent un jeu de caractéristiques, plus par habitude, que par une réelle analyse du problème. Ainsi, pour pallier ce manque de connaissances du domaine ou cette routine, nous proposons une méthode permettant de sélectionner automatiquement le jeu de caractéristiques le mieux adapté au problème considéré. Le module de reconnaissance mis en place repose sur un système d'inférence par règles floues automatiquement générée à partir d'un lot d'apprentissage. Choisir un jeu de caractéristiques quasi optimal doit permettre de réduire la complexité de la base de règles et ainsi garder une bonne interprétabilité du mécanisme de reconnaissance. Pour l'application testée, la méthode itérative implémentée a déterminé un choix de caractéristiques très similaire au choix de l'expert. Néanmoins, la différence entre les deux jeux de caractéristiques entraîne une différence non négligeable en terme de taux de reconnaissance. En effet, bien que le nombre de règles générées est plus important que celui obtenu avec le choix par expertise, la méthode itérative améliore de 4% les performances.

**Abstract** – This article presents an iterative feature selection method based on the Choquet integral and the typicality analysis. The proposed method is applied to wood defect recognition on wooden boards, domain in which few data is available in learning stage. The choice of a subset of features among these extracted from an image is not easy to obtain for a non-expert of the domain. With the use, it appears that even an expert of the domain chooses a set of features more by habits than by analysis of the problem. Thus, to lessen the impact of this lack of knowledge or this routine, we propose a method allowing to automatically select the most adapted set of features to the considered problem. The used recognition module relies on a rule-based system which is automatically generated from the learning database. Choosing an almost optimal set of features must reduce the rule base complexity and thus keeping a good interpretability of the recognition mechanism. For the tested application, the implemented iterative method determined a choice of feature close to the expert's one. Nevertheless, the difference between the two sets of features involves a significant difference in term of recognition rate. Indeed, although the number of generated rules is more important than the one obtained with the expertise choice, the iterative method allows to obtain a gain of 4% in term of recognition rate.

## 1. Introduction

Dans beaucoup d'applications de reconnaissance de formes, la sélection des caractéristiques est une tâche difficile. L'information à extraire de l'image n'est pas toujours triviale et pour être sûr d'avoir une quantité d'information suffisante, le nombre de caractéristiques calculées sur l'image peut rapidement croître. Hormis la

quantité d'information suffisante pour atteindre l'objectif fixé, il existe une réelle difficulté au niveau de la qualité de l'information à extraire. Dans cette optique, la réduction du fossé sémantique [1] semble être nécessaire afin de lier les besoins exprimés sous forme qualitative aux caractéristiques à extraire de l'image pour parvenir à l'objectif. La sélection des

caractéristiques répondant à ce critère revient alors à un problème de réduction de dimensionnalité. Il peut être décrit comme un problème d'optimisation où un sous-ensemble de caractéristiques est cherché dans le but de maximiser les performances de classification du système de reconnaissance. Le processus de sélection des caractéristiques les plus représentatives peut être vu sous deux angles, soit à partir de techniques de fouilles de données, pour lesquelles l'utilisation d'algorithmes génétiques est largement répandu [2-4], soit à partir de connaissances expertes du domaine concerné.

La méthode de reconnaissance utilisée pour l'application repose sur un système à base de règles linguistiques floues [5] générées automatiquement à partir d'un lot d'apprentissage [6] et dont la complexité dépend en partie du nombre de caractéristiques utilisés (l'autre facteur d'influence étant le nombre de termes de flouification pour chaque caractéristique). Jusqu'à présent, la sélection des caractéristiques est effectuée par exploitation des connaissances liées aux deux domaines d'expertise mis en jeu pour la reconnaissance des défauts du bois. Ces deux domaines complémentaires concernent le domaine du bois (définition des défauts permettant une reconnaissance à l'œil nu) et le domaine de la vision (définition des caractéristiques du système de vision qui peuvent être utilisés pour reconnaître les défauts). Le problème d'une telle sélection réside en sa validation assurant, d'une part, que les caractéristiques choisies sont bien les meilleures, et que, d'autre part, elles portent bien toute l'information nécessaire pour effectuer une reconnaissance correcte. Le contexte d'application de cette étude implique que la sélection de caractéristiques doit être faite avec un lot d'apprentissage contenant peu de données. La réduction de la dimensionnalité du problème doit également permettre de garder une interprétabilité à échelle humaine des règles générées. En d'autres termes, la base de règles doit pouvoir être facilement interprétée par un utilisateur humain.

## 2. Méthode de sélection de caractéristiques proposée

Beaucoup de systèmes de combinaison ont été proposés et comparés dans la littérature [7-9]. La méthode que nous proposons consiste à analyser le lot de données d'apprentissage utilisé pour générer la base de règles. L'extraction des sous-ensembles de caractéristiques utilise l'analyse des typicalités [10] et les intégrales de Choquet [11-14]. L'analyse des typicalités est fondée sur la ressemblance et la dissemblance d'un échantillon par rapport aux classes composant le lot d'apprentissage, permettant ainsi de déterminer, pour chaque paramètre, les classes typiquement représentées par ce paramètre. L'intégrale de Choquet fait partie des techniques d'agrégation basée sur les intégrales floues. Les intégrales floues, et l'intégrale de Choquet en particulier, ont été utilisées avec succès comme opérateurs de fusion dans beaucoup d'applications [15-17].

Le processus global de sélection des caractéristiques est itératif et peut être divisé en trois étapes, dont une correspond à la phase d'initialisation. La figure 1 illustre le processus de sélection que nous proposons.

- **Etape 0.** Appliquer l'analyse des typicalités pour choisir un premier sous-ensemble de caractéristiques et valider ce choix en entraînant et en testant le moteur d'inférence flou avec ces caractéristiques. Un taux de reconnaissance globale est obtenu et est considéré comme le taux de reconnaissance initial (TRI).
- **Etape 1.** A partir de ce premier sous-ensemble de caractéristiques, appliquer le processus de sélection basé sur l'intégrale de Choquet pour déterminer les  $k$  caractéristiques les moins représentatives.
- **Etape 2.** Générer le modèle de reconnaissance en supprimant de la liste initiale (étape 0) sans la première caractéristique la moins représentative et tester ce sous-ensemble. Le taux de reconnaissance obtenu est mémorisé (MTR). Répéter le processus pour les  $k-1$  caractéristiques suivantes.

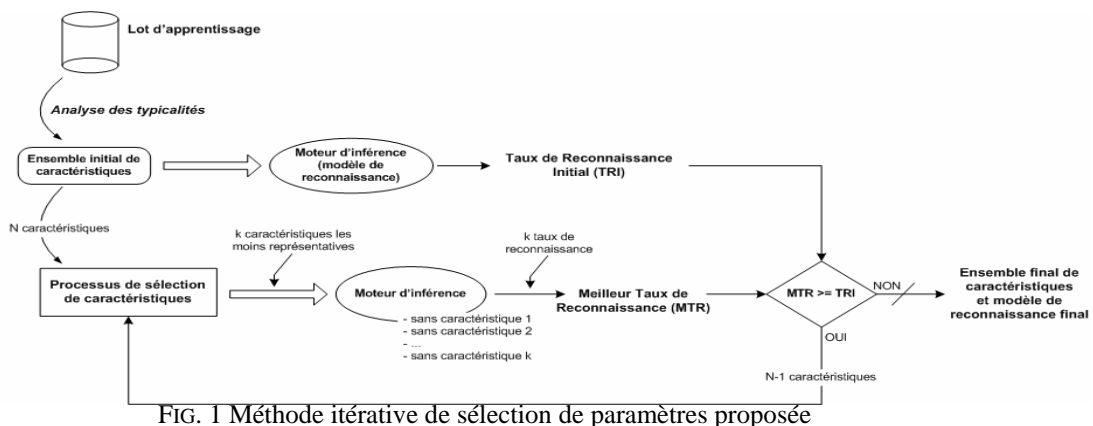


FIG. 1 Méthode itérative de sélection de paramètres proposée

Si un des taux obtenus dans l'étape 2 est meilleur ou égale à celui obtenu à l'étape 0 ( $MTR \geq TRI$ ) alors enlever la caractéristique la moins représentative associée à ce taux et recommencer le processus de sélection avec  $k-1$  caractéristiques. Sinon, garder le dernier sous-ensemble donnant un  $MTR \geq TRI$  et le prendre pour sous-ensemble final lors de la génération du modèle.

### 3. Inférence Floue

Le choix d'une méthode basée sur la logique floue, pour notre application de détection des défauts du bois, peut être justifié par deux raisons. Premièrement, les défauts à reconnaître sont intrinsèquement flous (transition graduelle entre le bois sain et les défauts). Les caractéristiques extraites des images sont alors incertaines (mais précisément calculées) et l'utilisation de la logique permet alors de prendre cette incertitude en compte. Deuxièmement, le client exprime ses besoins sous une forme linguistique ; les classes de sortie sont alors subjectives et souvent non disjointes (frontières non strictes entre une classe de défauts de type petit nœud et une classe de défauts représentant un nœud, par exemple).

Le mécanisme d'inférence implémenté utilise un système à base de règles linguistiques floues où chaque règle est automatiquement générée à partir d'un lot d'apprentissage. Parmi les différentes méthodes de génération de bases de règles, notre choix s'est porté sur l'algorithme de Ishibuchi-Nozaki-Tanaka [18]. Les règles utilisées sont des règles conjonctives dont l'inférence repose sur le modèle de Larsen. Les règles obtenues, sous forme d'une matrice, sont utilisées pour classifier les différents échantillons non étiquetés. La figure 2 représente, de manière succincte, le principe du « Fuzzy Reasoning Classifier » (FRC) développé autour de cet algorithme et pour notre application, soulignant les étapes d'apprentissage et d'exploitation du modèle [5].

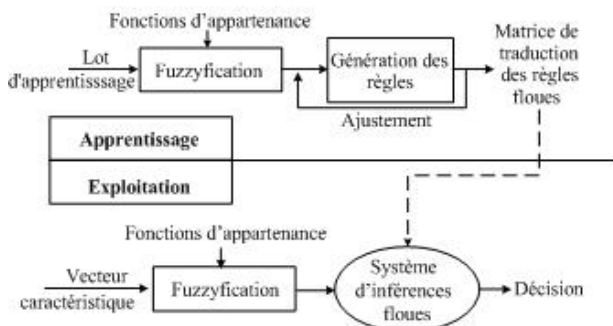


FIG. 2 – Principe du mécanisme d'inférence

### 4. Résultats expérimentaux

Les résultats ont été obtenus à partir d'une base de données composée de 877 échantillons répartis en 9

classes de défauts. Le lot de données d'apprentissage sur laquelle la méthode de sélection est appliquée, est composé de 250 échantillons. Le lot utilisé pour tester et valider les sous-ensembles construits est donc composé de 627 échantillons. Le système d'inférences floues est une inférence unique où toutes les caractéristiques sont en entrée du modèle et toutes les classes à reconnaître en sortie.

Initialement, la sélection des caractéristiques est faite par expertise. Dans le cas présent, l'expert a choisi 6 caractéristiques (parmi les 19 extraites des images par le système de vision industriel du client) classiquement utilisées pour reconnaître les défauts présents dans les images de la base de données. Le taux de reconnaissance obtenu est de 74,2% et 242 règles sont générées. Dans la méthode que nous proposons, la première étape est l'utilisation des typicalités pour choisir le premier sous-ensemble de caractéristiques.

La méthode donne 9 caractéristiques significatives pour reconnaître toutes les classes. Parmi ces caractéristiques, 4 représentent une notion de forme, 1 représente une notion de taille et 4 représentent une notion de couleur. Le résultat de cette analyse est double : le taux de reconnaissance augmente sensiblement pour atteindre 77,8%, mais 4641 règles sont générées rendant très difficile l'interprétation du modèle. L'intérêt de la partie itérative est de décroître le nombre de caractéristiques à utiliser. L'objectif est donc de diminuer la complexité du modèle sans diminuer le taux de reconnaissance obtenu avec l'analyse des typicalités. A chaque étape de la partie itérative, notre méthode de sélection fournit une analyse complète du sous-ensemble courant, sous forme matricielle, indiquant les interactions entre les caractéristiques ainsi que l'importance de chaque caractéristique.

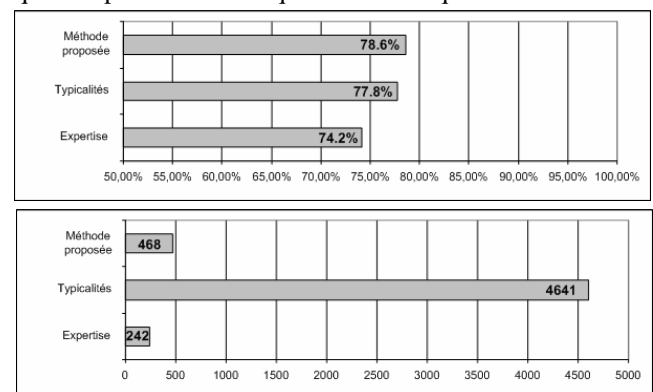


FIG. 3 – Taux de reconnaissance et nombre de règles pour les trois méthodes de sélection

Dans notre expérimentation, la partie itérative conduit à supprimer trois caractéristiques du sous-ensemble initial (fourni par les typicalités) permettant ainsi d'augmenter le taux de reconnaissance à 78,6% (+ 0,8%) et de diminuer la complexité du modèle à 468 règles. La figure 3 résume les résultats obtenus avec les trois méthodes de sélection.

## 5. Conclusion

La méthode de sélection itérative proposée dans cet article est fondée sur une analyse du lot d'apprentissage en trois étapes : la première, représentant l'initialisation, permet de choisir un premier sous-ensemble de caractéristiques à partir d'une analyse des typicalités. Les deuxième et troisième étapes forment la partie itérative de la méthode qui permet de réduire la dimension du problème tout en gardant un taux de reconnaissance élevé. L'avantage d'une telle méthode repose sur la potentialité à être efficace avec un lot d'apprentissage réduit (comportant peu d'échantillons par classe) afin d'effectuer une sélection de caractéristiques proche de l'optimum. Enfin, cette diminution du nombre de caractéristiques intervient dans un contexte où l'expertise est forte (besoins client et connaissances utiles sur le système de vision) et donc où l'interprétabilité de la base de règle est primordiale. En effet, plus la base de règles sera de taille humaine (compréhensible et assimilable par un être humain), meilleure sera son interprétabilité afin de déterminer le comportement du système de reconnaissance en situation de résolution du problème considéré, à savoir nommer correctement un objet inconnu.

Les résultats expérimentaux montrent que la méthode proposée permet de choisir un sous-ensemble de caractéristiques plus discriminants, augmentant le taux de reconnaissance en comparaison du choix effectué par expertise tout en conservant encore un certain degré d'interprétabilité du modèle. Nos travaux futurs viseront à réduire le nombre de règles générées. Pour cela, nous pensons étendre la méthode proposée dans cet article à l'extraction des caractéristiques pertinentes pour chaque classe et non pour l'ensemble des classes, comme effectué actuellement. Cela aboutira à une structure de modèle différente reposant sur une structure arborescente. Cette structure est basée sur un enchaînement hiérarchique de plusieurs inférences uniques dont les sorties permettent une séparation plus nette des classes de défauts à reconnaître (par exemple, séparation des défauts plutôt ronds des défauts plutôt allongés). L'objectif étant alors de diminuer le nombre de règles générées et ainsi d'améliorer l'interprétabilité du mécanisme de reconnaissance.

## Références

- [1] C. HUDELLOT, J. ATIF, I. BLOCH. Ontologie de relations spatiales floues pour l'interprétation d'images. *LFA 2006, 14<sup>ème</sup> rencontres francophones sur la Logique Floues et ses Applications*, Toulouse, France, p.363-370, 2006.
- [2] H. HANDELS, TH. ROß, J. KREUSCH, H.H. WOLFF, S.J. PÖPPL. Feature selection for optimized skin tumor recognition using genetic algorithms. *Artificial Intelligence in Medicine*, vol. 16, p. 283-297, 1999.
- [3] N.K. JAIN, V.K. JAIN, K. DEB. Optimization of process parameters of mechanical type advanced machining processes using genetic algorithms. *International Journal of Machine Tools & Manufacture*, vol. 47, p. 900-919, 2007.
- [4] C.H. HUANG, C-J. WANG. A GA-based feature selection and parameters optimization for support vector machines. *Expert Systems with Applications*, vol. 31, p. 231-240, 2006.
- [5] D. DUBOIS, H. PRADE. What are Fuzzy rules and how to use them?. *Fuzzy Sets and Systems*, vol. 84, p. 169-185, 1996.
- [6] E. SCHMITT, C. MAZAUD, V. BOMBARDIER, P. LHOSTE. A Fuzzy Reasoning Classification Method for Pattern Recognition. *15th International Conference on Fuzzy Systems, FUZZIEEE'06*, Vancouver, Canada, p. 5998-6005, 2006.
- [7] J. KITTLER, M. HATEF, R. DUIN, AND J. MATAS. On combining classifiers. *IEEE Transactions on PAMI*, vol. 20(3), p. 226-239, 1998.
- [8] L.I. KUNCHEVA, C.J. WHITAKER. Measures of diversity in classifier ensembles. *Machine Learning*, vol. 51, p. 181-207, 2003.
- [9] D. RUTA, B. GABRYS. An overview of classifier fusion methods. *Computing and Information System*, vol. 7, p. 1-10, 2000.
- [10] J. FOREST, M. RIFQI, B. BOUCHON-MEUNIER. Segmentation de classes pour l'amélioration de la construction de prototypes flous : visualisation et caractérisation de classes non homogènes. *Rencontres Francophones sur la Logique Floue et ses Applications LFA*, Toulouse, France, p. 29-36, 2006.
- [11] G. CHOQUET. Theories of capacities. *Annales de l'institut Fourier*, vol. 5, p. 131-295, 1953.
- [12] M. GRABICSH, J.M. NICOLAS. Classification by fuzzy integral – performance and tests. *Fuzzy Sets and Systems, Special Issue on Pattern recognition*, vol. 65, p. 255-271, 1994.
- [13] J.L. MARICHAL. Aggregation of interacting criteria by means of the discrete Choquet integral. In *Aggregation operators: new trends and applications*, p. 224-244, Physica-Verlag GmbH, Heidelberg, Allemagne, 2002.
- [14] T. MUROFUSHI, M. SUGENO. A theory of fuzzy measures: representations, the Choquet integral, and null sets. *Journal of Mathematical Analysis and Applications*, vol. 159, p. 532-549, 1991.
- [15] S. B. CHO, J. H. KIM. Combining multiple neural networks by fuzzy integral for robust classification. *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 25(2), p. 380-384, February 1995.
- [16] H. TAHANI, J.M. KELLER. Information fusion in computer vision using the fuzzy integral. *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 20(3), May/June 1990.
- [17] Y. S. CHOI, D. KIM. Relevance feedback for content-based image retrieval using the choquet integral. *IEEE trans. on Knowledge and Data Engineering*, vol. 16(10), p. 1185-1199, 2004.
- [18] H. ISHIBUCHI, K. NOZAKI, H. TANAKA. A simple but powerful heuristic method for generating fuzzy rules from numerical data. *Fuzzy Sets and Systems*, vol. 86, p. 251-270, 1997.