

Fusion multi-vues à partir de fonctions de croyance pour la classification d'objets

Hicham Laanaya*, Arnaud Martin*

*ENSIETA, E³I²-EA3876, 2 rue François Verny, 29806 Brest Cedex 9
{Hicham.Laanaya, Arnaud.Martin}@ensieta.fr,
<http://www.ensieta.fr/e3i2/>

Résumé. Nous présentons dans cet article une approche de fusion crédibiliste pour la classification d'objets à partir d'images multi-vues. Cette approche est appliquée sur des données générées à partir de trois formes de base (un cercle, un hexagone et un octogone). Ces données simulent une sortie de classifieur et représentent ainsi les positions des objets. Les résultats obtenus montrent l'intérêt de l'exploitation de l'information extraite sur plusieurs vues pour la classification d'un même objet.

1 Introduction

Cet article présente une approche de classification d'objets fondée sur l'utilisation des informations extraites à partir d'images de l'objet sous plusieurs vues. Sur chaque vue on extrait des informations (attributs) qu'on utilise pour la classification de cet objet, ainsi, chaque vue, après la phase de classification, donne une information sur la classe et la position de l'objet. Une fusion de ces classes paraît nécessaire pour augmenter le taux de reconnaissance de l'objet (*cf.* figure 1).

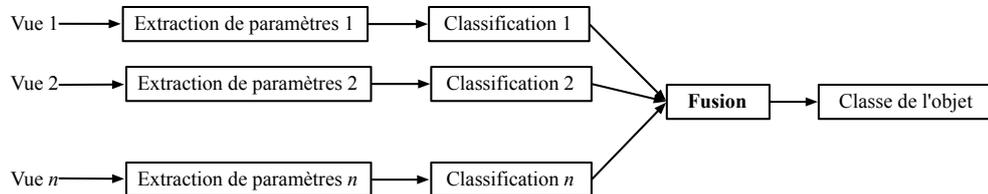


FIG. 1 – Fusion multi-vues pour la classification d'un objet

La fusion multi-vues pour la classification d'objets peut donc être vue comme un problème de fusion de résultats de classification de chaque vue (ou fusion de classifieurs). Les méthodes classiques issues de la théorie de l'incertain (fusion par vote, fusion par vote pondéré, fusion bayésienne, fusion par les fonctions de croyance, ...) peuvent donc être utilisées pour réaliser cette fusion (*cf.* Laanaya et al. (2008), Martin (2005a)). Dans ces théories de l'incertain deux notions bien distinctes sont essentielles afin de bien modéliser les imperfections des données : l'incertitude et l'imprécision. L'incertitude caractérise un degré de conformité à la réalité (défaut qualitatif de l'information), tandis que l'imprécision mesure un défaut quantitatif de l'information (par exemple une erreur de mesure).

Plusieurs études dans la littérature utilisent l'information extraite sur plusieurs vues pour la détection d'objets, par exemple dans la thèse de Daniel (1998), ils ont étudié l'apport de l'utilisation de plusieurs vues sonar pour améliorer les taux de classification des objets. Dans le même contexte, Quidu (2001); Aridgides et al. (2001) utilisent la classification multi-vues pour la détection des mines. En imagerie sonar, la classification multi-vues a plusieurs intérêts tels que le positionnement d'un robot autonome, et nécessite généralement une étape de recalage des images Dhibi et al. (2008). La fusion multi-vues a été aussi utilisée pour la classification des sexes en se fondant sur une séquence de marche Huang et Wang (2007). Dans le domaine de la télédétection, Milisavljevic et al. (2008) ont utilisé une approche fondée sur la théorie des fonctions de croyance et la théorie des possibilités pour la détection de mines antipersonnel.

Cet article présente une étude théorique de la fusion pour la classification multi-vues qui a pour objectif de montrer la faisabilité et l'intérêt de la combinaison des informations issues de plusieurs classifieurs à partir d'images prises sous différents angles, en particulier en imagerie sonar. Le manque de ce type de données nous a poussé à réaliser cette étude préliminaire à partir de données générées.

Notre choix s'est porté sur la théorie des fonctions de croyance qui permet, dans un même cadre théorique, de bien modéliser l'incertitude et l'imprécision et offre des avantages également pour modéliser le manque d'information selon le point de vue.

L'organisation de cet article est la suivante. Dans la section 2, nous présentons les bases théoriques des fonctions de croyance. Nous donnerons ensuite une description des données générées et utilisées pour la validation de l'approche de fusion multi-vues. Les résultats de classification sont donnés dans la section 3.2.

2 Fusion pour la classification multi-vues

La fusion pour la classification multi-vues exploite l'information acquise sur un objet pour plusieurs vues. Cette information, dans la plupart du temps, est entachée d'imperfections liées généralement au milieu étudié et aux capteurs utilisés. Les fonctions de croyance présentées dans la section 2.1 permettent une meilleure modélisation de ce genre d'information.

Le principe général d'une approche de fusion d'informations est décrit par la figure 2 Martin (2005b). Ainsi nous considérons les informations issues des différents classificateurs à fusionner, ainsi que des informations supplémentaires et des connaissances externes liées à l'application pour aider la combinaison.

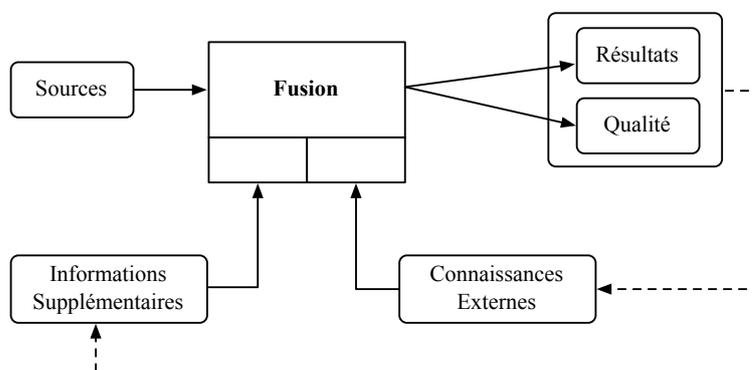


FIG. 2 – Représentation de la fusion

Le processus de la fusion d'information est décrit par quatre étapes : la modélisation, l'estimation, la combinaison et la décision (cf. figure 3). La modélisation définit le choix du formalisme, qui sera dans notre cas la théorie des fonctions de croyance. L'estimation permet de définir les fonctions choisies dans l'étape de modélisation en fonction de l'application. La combinaison est la phase de regroupement des informations. La dernière étape consiste à prendre la décision sur le résultat de la combinaison.

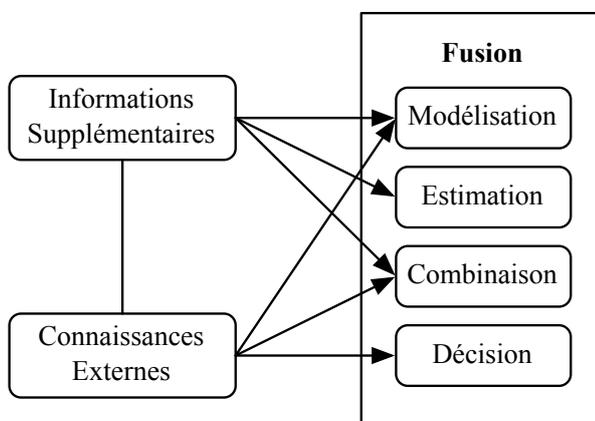


FIG. 3 – Représentation du nœud de fusion

2.1 Fonctions de croyance

Nous proposons ici l'utilisation de la théorie des fonctions de croyance pour la fusion multi-vues d'objets en vue d'une meilleure classification.

La théorie des fonctions de croyance est fondée sur la manipulation des fonctions de masse. Les fonctions de masse sont définies sur l'ensemble de toutes les disjonctions du cadre de discernement $\Theta = \{C_1, \dots, C_N\}$ et à valeurs dans $[0, 1]$, où C_q représente l'hypothèse "l'observation appartient à la classe q ". Généralement, il est ajouté une condition de normalité, donnée par :

$$\sum_{A \in 2^\Theta} m(A) = 1, \quad (1)$$

où $m(\cdot)$ représente la fonction de masse. La première difficulté est donc de définir ces fonctions de masse selon le problème. À partir de ces fonctions de masse, d'autres fonctions de croyance peuvent être définies, telles que les fonctions de crédibilité, représentant l'intensité que toutes les sources croient en un élément, et telles que les fonctions de plausibilité représentant l'intensité avec laquelle on ne doute pas en un élément.

De façon à estimer les fonctions de masse à combiner, Appriou (2002) propose deux modèles répondant à trois axiomes qui impliquent la considération de N fonctions de masse aux seuls éléments focaux possibles $\{C_q\}$, $\{C_q^c\}$ et Θ . Un axiome garantit de plus l'équivalence avec l'approche bayésienne dans le cas où la réalité est parfaitement connue (méthode optimale dans ce cas). Ces deux modèles sont sensiblement équivalents sur nos données, nous utilisons dans cet article le modèle donné par :

$$\begin{cases} m_{iq}(C_q)(x) &= \frac{\alpha_{iq} R_i p(V_i(x)/C_q)}{1 + R_i p(V_i(x)/C_q)} \\ m_{iq}(C_q^c)(x) &= \frac{\alpha_{iq} R_i}{1 + R_i p(V_i(x)/C_q)} \\ m_{iq}(\Theta)(x) &= 1 - \alpha_{iq} \end{cases} \quad (2)$$

où p est une probabilité, $R_i = (\max_{i,q} p(V_i(x)/C_q))^{-1}$ est un facteur de normalisation, et $\alpha_{iq} \in [0, 1]$ est un coefficient d'affaiblissement permettant de tenir compte de la fiabilité de l'information fournie par la vue $i : V_i(x)$ pour une classe C_q , que nous choisissons ici égale à 0.95. La difficulté de ce modèle est alors l'estimation des probabilités $p(V_i(x)/C_q)$. Dans le cas où la donnée $V_i(x)$ de la vue i est la réponse d'un classifieur exprimée sous la forme de la classe (donnée symbolique), l'estimation de ces probabilités peut être faite par les matrices de confusion sur une base d'apprentissage.

La combinaison des N (nombre de vues dans cette étude) fonctions de masse que nous employons ici est la combinaison conjonctive non normalisée proposée par Smets (1990a) définie pour deux fonctions de masse m_1 et m_2 et pour tout $A \in 2^\Theta$ par :

$$m(A) = (m_1 \oplus m_2)(A) = \sum_{B \cap C = A} m_1(B) m_2(C). \quad (3)$$

De nombreuses autres règles ont été proposées, un bref état de l'art ainsi que de nouvelles règles de combinaison gérant le conflit sont données par Martin et Osswald (2007).

Afin de conserver un maximum d'informations, il est préférable de rester à un niveau crédal (*i.e.* de manipuler des fonctions de masse) pendant l'étape de combinaison des informations pour prendre la décision sur les fonctions de masse issues de la combinaison. Si la décision prise par le maximum de crédibilité peut être trop pessimiste, la décision issue du maximum de plausibilité est bien souvent trop optimiste. Le maximum de la probabilité pignistique, introduite par Smets (1990b), reste le compromis le plus employé. La probabilité pignistique est donnée pour tout $X \in 2^\Theta$, avec $X \neq \emptyset$ par :

$$\text{betP}(X) = \sum_{Y \in 2^\Theta, Y \neq \emptyset} \frac{|X \cap Y|}{|Y|} \frac{m(Y)}{1 - m(\emptyset)}. \quad (4)$$

3 Expérimentations

Nous présentons dans cette partie une description des données générées utilisées pour la validation de l'approche de fusion crédibiliste pour la classification multi-vues. Nous donnons dans la section 3.2 les résultats obtenus pour différentes configurations : on utilise un nombre différent de vues et un niveau de confusion différent entre les objets considérés.

3.1 Données utilisées

Nous avons utilisé pour la génération de notre base de données trois formes de base : un cercle (forme invariante par rotation), un hexagone et un octogone vues à 0° avec $\sigma = 0$.

Ces formes ont été considérées sur $N = 36$ vues, de 0° à 350° avec un pas de 10° , en ajoutant un bruit au niveau des bords des formes. Ce bruit est ajouté aléatoirement au niveau des bords en utilisant une gaussienne multipliée par la fonction de la forme du bruit utilisée, représentée dans la figure 4.

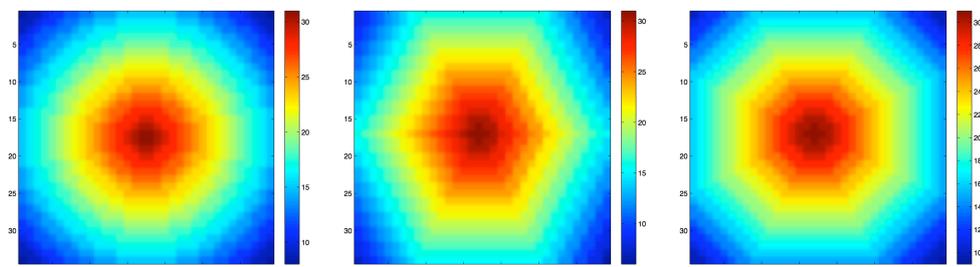


FIG. 4 – Les fonctions utilisées pour brouter les formes de base

La base de données est constituée d'images binaires de taille 128×128 pixels, la forme de base se situe au centre de cette image et de taille 32×32 pixels. La figure 5 donne un exemple de quelques images de la base de données. Le tableau 1 donne quelques statistiques sur la base de données où la classe 1 désigne la classe des cercles, la classe 2 celle des hexagones et la classe 3 la classe des octogones.

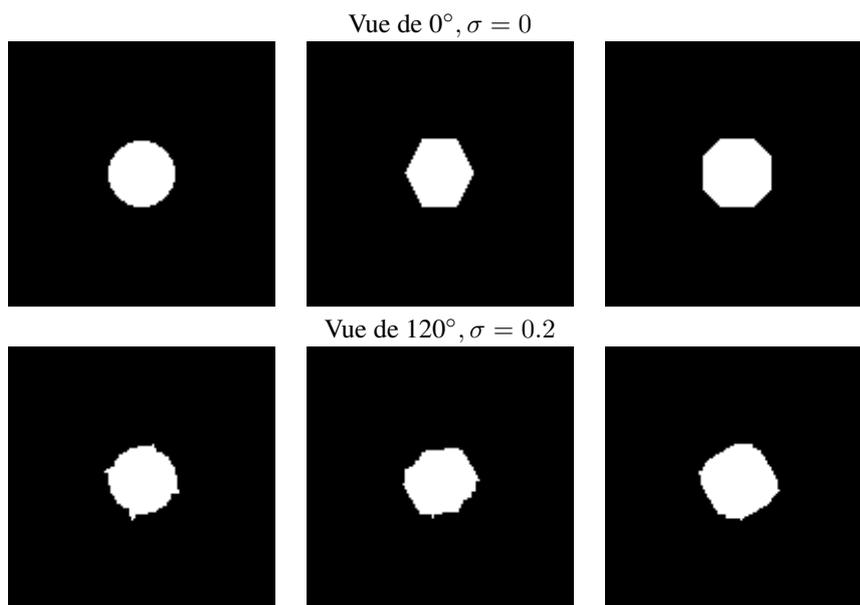


FIG. 5 – Trois images de la base de données utilisée

Classes	{Cercle, Hexagone, Octogone}
Nombre de classe	3
Effectif	1296 (432 pour chaque classe)
Vues	de 0° à 350° avec un pas de 10°
Certitudes	de 0.1 à 0.9 avec un pas de 0.1

TAB. 1 – Quelques statistiques sur la base de données générées

Chaque objet o est caractérisé par un vecteur $V_o = (V_o^i, i = 1, 2, 3, 4)$: chaque élément de ce vecteur représente le nombre de pixels qui entre en confusion avec les autres forme de base.

Nous avons considéré l'approche d'Appriou (2002) pour l'estimation des fonctions de masse à partir des paramètres calculés sur les objets. Ces fonctions sont estimées en utilisant une matrice de confusion sur les données d'apprentissage. Nous avons considéré deux approches pour le calcul de cette matrice de confusion : une première approche fondée sur les paramètres calculés directement sur les objets (vecteurs V_o) (La matrice de confusion est notée $MCnoSVM$) et une deuxième approche qui utilise une classification par SVM Vapnik (1998) sur la base d'apprentissage (matrice de confusion notée $MCSVM$). La base d'apprentissage est constituée de 648 images (216 images pour chaque classe) tirées aléatoirement en considérant toutes les certitudes.

3.2 Résultats

Nous donnons dans cette section les résultats obtenus pour la classification des objets en utilisant la fusion multi-vues. Pour cela, nous avons utilisé différentes valeurs pour le paramètre (σ) qui contrôle le bruit au bord des formes de base. Nous avons utilisé des valeurs entre 0 et 1 pour σ . Les résultats sont obtenus en utilisant les matrices de confusion normalisées présentées dans le tableau 2. Ces matrices ont été calculées en utilisant des bases d'apprentissage (de 648 images) tirées aléatoirement. Nous avons utilisé les SVM pour la classification en utilisant le logiciel *libSVM* développé par Chang et Lin (2001).

σ	$MCnoSVM$ (%)	$MCSVM$ (%)
0.4	$\begin{pmatrix} 81.81 & 12.50 & 5.68 \\ 13.19 & 79.27 & 7.54 \\ 12.10 & 11.05 & 76.85 \end{pmatrix}$	$\begin{pmatrix} 71.76 & 25.93 & 2.31 \\ 25.00 & 72.22 & 2.78 \\ 0.93 & 0.46 & 98.61 \end{pmatrix}$
0.5	$\begin{pmatrix} 77.35 & 16.32 & 6.34 \\ 21.86 & 69.01 & 9.12 \\ 9.98 & 17.11 & 72.91 \end{pmatrix}$	$\begin{pmatrix} 65.74 & 31.02 & 3.24 \\ 32.87 & 63.43 & 3.70 \\ 3.24 & 1.85 & 94.91 \end{pmatrix}$
0.6	$\begin{pmatrix} 72.61 & 17.79 & 9.60 \\ 11.26 & 75.11 & 13.62 \\ 16.03 & 15.37 & 68.60 \end{pmatrix}$	$\begin{pmatrix} 61.11 & 31.02 & 7.87 \\ 34.72 & 60.19 & 5.09 \\ 3.70 & 4.17 & 92.13 \end{pmatrix}$
0.7	$\begin{pmatrix} 84.32 & 10.65 & 5.04 \\ 5.75 & 90.72 & 3.53 \\ 7.60 & 16.33 & 76.06 \end{pmatrix}$	$\begin{pmatrix} 59.26 & 33.33 & 7.41 \\ 34.72 & 56.02 & 9.26 \\ 6.48 & 4.63 & 88.89 \end{pmatrix}$
0.8	$\begin{pmatrix} 91.00 & 2.47 & 6.52 \\ 6.17 & 78.76 & 15.07 \\ 10.24 & 13.11 & 76.65 \end{pmatrix}$	$\begin{pmatrix} 59.72 & 31.02 & 9.26 \\ 30.56 & 55.56 & 13.89 \\ 12.50 & 8.80 & 78.70 \end{pmatrix}$
0.9	$\begin{pmatrix} 88.12 & 2.26 & 9.62 \\ 10.49 & 83.60 & 5.91 \\ 12.91 & 30.60 & 56.49 \end{pmatrix}$	$\begin{pmatrix} 60.19 & 28.70 & 11.11 \\ 34.26 & 55.09 & 10.65 \\ 11.57 & 9.72 & 78.70 \end{pmatrix}$

TAB. 2 – Matrices de confusion calculées sur les données d'apprentissage et pour chaque valeur de σ

Les matrices de confusion normalisées pour le cas de la classification avant l'étape de la fusion sont données dans le tableau 3. Nous remarquons que le taux de classification décroît en fonction de σ : plus on a une superposition entre les classes, exprimée par l'utilisation de σ , plus le taux de classification est faible.

La figure 6 donne les résultats pour la fusion en utilisant une matrice de confusion calculée directement sur les données d'apprentissage ($MCnoSVM$) et une matrice de confusion calculée avec un classifieur de base ($MCSVM$) (SVM avec un noyau linéaire dans cette étude). Nous avons comparé les résultats par une approche classique de fusion par vote et par l'approche crédibiliste que nous avons décrit dans la section 2.1. Chaque courbe de la figure 6 représente la variation du taux de classification en fonction du nombre de vues. Notons que les vues sont tirées aléatoirement et que nous avons effectué plusieurs $10 \log_e(A_N^p)$ tirages aléatoires pour un nombre de vues égale à p . Nous n'avons pas effectué toutes les possibilités pour chaque nombre de vues, puisque le nombre du choix de vues possibles est égale à A_N^p (où $N = 36$ est le nombre de vues maximal et p le nombre de vues à choisir. Exemple : $A_{36}^{10} = 922.393.263.052.800$). Nous avons divisé la base de données en deux bases, une pour l'apprentissage du classifieur (et aussi pour le calcul des matrices de confusion utilisées pour l'estimation des fonctions de masse) et une base pour le test. Ces deux bases de données contiennent le même nombre de données avec le même nombre d'images pour chacune des trois classes.

Nous remarquons que pour les différentes valeurs de σ , on arrive à dépasser le taux de classification obtenu sans fusion (cf. tableau 3) à partir de l'utilisation de 2, 3 ou 4 vues. Nous remarquons aussi qu'on atteint les 100% de taux de bonne classification pour $\sigma = 0.4, 0.5, 0.6$ et 0.8 . Les différents résultats de classification pour $\sigma = 0.4, 0.5, 0.6, 0.7, 0.8$ et 0.9 ne sont pas significativement différents.

Fusion multi-vues à partir de fonctions de croyance pour la classification d'objets

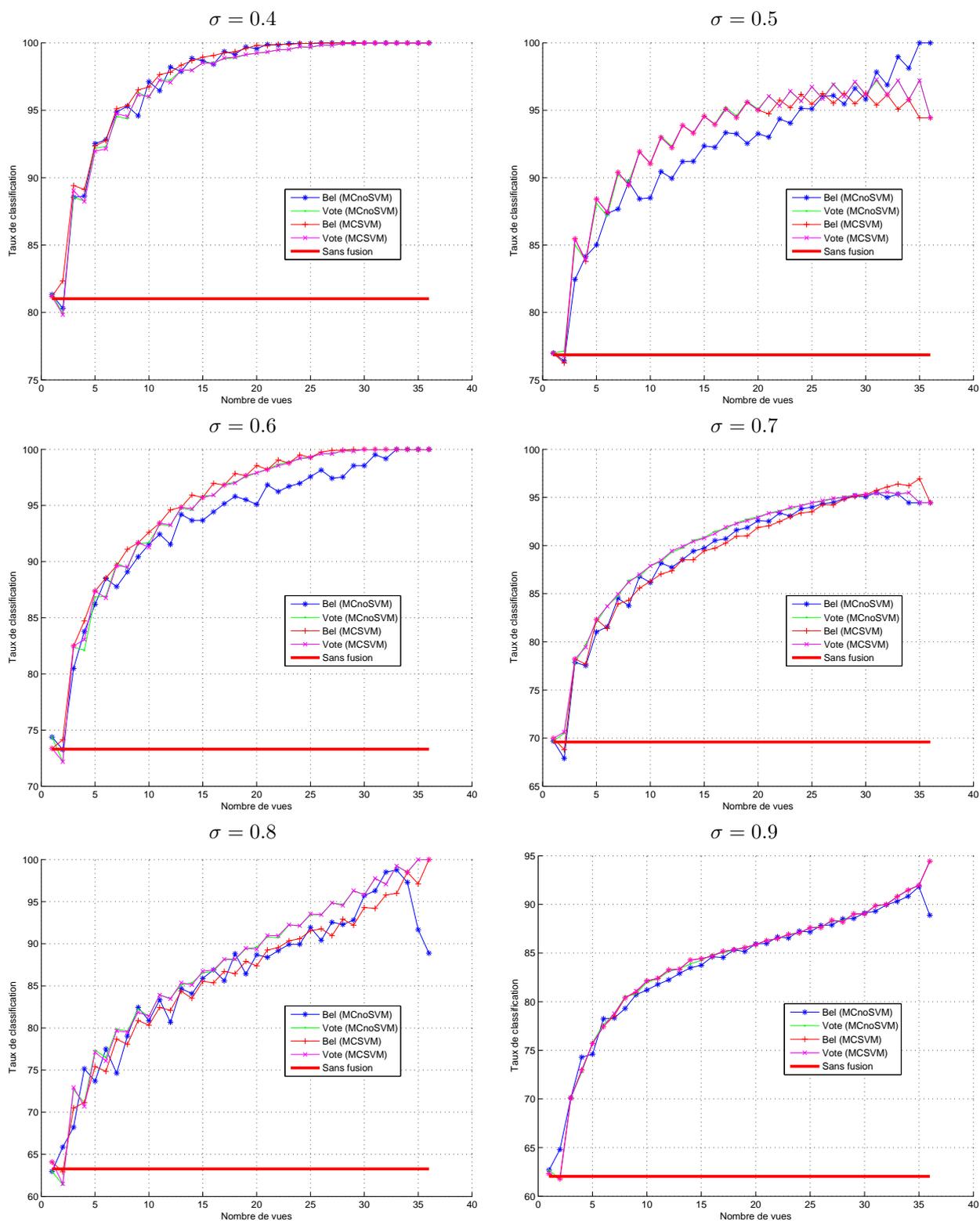


FIG. 6 – Résultats de fusion multi-vues en utilisant deux matrices de confusion différentes pour les deux approches de fusion, vote et crédibiliste

σ	Taux de classification (%)	Probabilités d'erreur (%)	Matrice de confusion (%)
0.4	81.02±3.20%	(20.14 20.49 2.08)	$\begin{pmatrix} 75.00 & 25.00 & 0.00 \\ 27.31 & 71.76 & 0.93 \\ 3.24 & 0.46 & 96.30 \end{pmatrix}$
0.5	76.85±3.40%	(26.62 22.57 2.89)	$\begin{pmatrix} 62.50 & 31.48 & 6.02 \\ 29.17 & 70.83 & 0.00 \\ 2.31 & 0.46 & 97.22 \end{pmatrix}$
0.6	73.30±3.54%	(26.50 26.04 7.52)	$\begin{pmatrix} 66.20 & 27.78 & 6.02 \\ 31.94 & 63.43 & 4.63 \\ 6.48 & 3.24 & 90.28 \end{pmatrix}$
0.7	69.60±3.65%	(31.13 28.24 9.03)	$\begin{pmatrix} 54.63 & 37.04 & 8.33 \\ 28.24 & 64.35 & 7.41 \\ 5.56 & 4.63 & 89.81 \end{pmatrix}$
0.8	63.27±3.78%	(35.42 32.87 14.35)	$\begin{pmatrix} 52.78 & 31.48 & 15.74 \\ 38.43 & 54.17 & 7.41 \\ 8.80 & 8.33 & 82.87 \end{pmatrix}$
0.9	62.04±3.80%	(34.95 34.38 16.09)	$\begin{pmatrix} 47.22 & 42.59 & 10.19 \\ 27.78 & 59.72 & 12.50 \\ 6.48 & 14.35 & 79.17 \end{pmatrix}$

TAB. 3 – Matrices de confusion normalisées avant la fusion multi-vues

4 Conclusion

Nous avons employé dans cet article l'étude de la fusion pour la classification multi-vues, appliquée à des données générées modélisant la sortie d'un classifieur tentant de discriminer trois types d'objets. Le but de la fusion est ici d'améliorer le taux de bonne classification en augmentant le nombre de vues. Nous avons quantifier l'amélioration en fonction du nombre de vues ainsi que du niveau de bruit sur les résultats fournis par le classifieur. L'approche étudiée, fondée sur les fonctions de croyance a permis une amélioration significative des résultats pour la classifications multi-vues. En effet nous avons pu avoir des taux de classification qui dépassent largement le taux de classification obtenu en utilisant uniquement une seule vue. Nous avons pu voir qu'on obtient des taux de classification de 100% à partir de l'utilisation de quelques vues pour des valeurs de σ assez petites (faible superposition entre les classes).

Les données générées ne modélisent que grossièrement les résultats possibles en sortie d'un classifieur. Nous essayerons d'appliquer cette approche sur des données réelles comme les images sonar pour la détection d'objets dans un milieu sous-marin (mines, épaves ...).

Références

- Appriou, A. (2002). *Décision et Reconnaissance des formes en signal*. Hermes Science Publication.
- Aridgides, T., M. F. Fernandez, et G. J. Dobeck (2001). Side-scan sonar imagery fusion for sea mine detection and classification in very shallow water. In A. C. Dubey, J. F. Harvey, J. T. Broach, et V. George (Eds.), *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, Volume 4394, pp. 1123–1134.
- Chang, C. C. et C. J. Lin (2001). Libsvm : a library for support vector machines. *Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>*.
- Daniel, S. (1998). Fusion multisource appliquée à la reconnaissance d'objets dans le milieu sous-marin. *Thèse de l'Université de Rennes 1, Rennes, FRANCE*.
- Dhibi, M., R. Courtis, et A. Martin (2008). Multi-segmentation of sonar images using belief function theory. In *Acoustics'08/ECUA08*, Paris, France.
- Huang, G. et Y. Wang (2007). Gender classification based on fusion of multi-view gait sequences. In *ACCV (1)*, pp. 462–471.
- Laanaya, H., A. Martin, D. Aboutajdine, et A. Khenchaf (2008). Classifier fusion for post-classification of textured images. *Information Fusion, 30 June-3 July*.
- Martin, A. (2005a). Fusion de classifieurs pour la classification d'images sonar. *RNTI Extraction des connaissances : Etat et perspectives*, 259–268.

- Martin, A. (Janvier 2005b). La fusion d'informations. *Polycopié de cours ENSIETA - Réf. : 1484, 117 pages.*
- Martin, A. et C. Osswald (2007). Toward a combination rule to deal with partial conflict and specificity in belief functions theory. In *International Conference on Information Fusion*, Québec, Canada.
- Milisavljevic, N., I. Bloch, et M. Acheroy (2008). *Multi-Sensor Data Fusion Based on Belief Functions and Possibility Theory : Close Range Antipersonnel Mine Detection and Remote Sensing Mined Area Reduction*, Chapter 4, pp. 392–418. Vienna, Austria : ARS I-Tech Education and Publishing.
- Quidu, I. (2001). Classification multi-vues d'un objet immergé à partir d'images sonar et de son ombre portée sur le fond. *Thèse de l'Université de Bretagne Occidentale, Brest, FRANCE.*
- Smets, P. (1990a). The Combination of Evidence in the Transferable Belief Model. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 12(5), 447–458.
- Smets, P. (1990b). Constructing the pignistic probability function in a context of uncertainty. *Uncertainty in Artificial Intelligence* 5, 29–39.
- Vapnik, V. N. (1998). *Statistical Learning Theory*. John Wesley and Sons.

Summary

We present in this paper a belief fusion approach for pattern recognition of multi-view images. This approach is applied on generated data from three basic shapes (circle, hexagon and octagon). These data represent the object position after a classifier algorithm. The results show the interest of the use of information extracted on multiple views for the classification of the same object.