Sacs de relations spatiales et de formes pour la reconnaissance d'images de scènes naturelles

Michaël Clément, Camille Kurtz, Laurent Wendling

Sacs de relations spatiales et de formes pour la reconnaissance d'images de scènes naturelles

Michaël Clément Camille Kurtz Laurent Wendling
Université Paris Descartes, LIPADE, équipe SIP - 45 rue des Saints-Pères, 75006 Paris, France
michael.clement@parisdescartes.fr

Résumé

Nous présentons une approche de type sacs de caractéristiques, fondée sur des descripteurs de position relative, pour la reconnaissance d'images. D'une part, étant donnée une décomposition initiale de l'image en un ensemble d'objets sémantiques, une description à partir de la Décomposition en Histogrammes de Forces (FHD) est introduite, modélisant à la fois la forme et les relations spatiales entre les objets composant l'image. D'autre part, une méthodologie originale d'apprentissage est présentée, afin de construire un vocabulaire homogène de formes et de configurations spatiales pour des tâches de classification d'images. Un avantage de cette stratégie est sa compatibilité avec les approches par sacs de caractéristiques classiques, permettant une représentation hybride d'informations locales et structurelles. Les résultats de classification obtenus sur un jeu de données d'images de scènes naturelles montrent l'intérêt de cette approche.

Mots Clef

Sacs de relations spatiales, descripteurs de position relative, histogrammes de forces, reconnaissance de scènes

Abstract

We present a bags of features approach for image recognition, based on relative position descriptors, modeling both the shape and spatial relationships between the objects composing the image. On the one hand, given an initial decomposition of the image into a set of semantic objects, a description using the Force Histograms Decomposition (FHD) is introduced. On the other hand, an original learning methodology is presented, in order to construct a homogeneous vocabulary of shapes and spatial configurations for image classification tasks. An advantage of this strategy relies on its compatibility with the more classical bags-of-features models, allowing a hybrid representation of structural and local information. The results of classification obtained on a dataset of natural scenes images highlight the interest of this approach.

Keywords

Bags of relations, relative position descriptors, force histograms, scene recognition

1 Introduction

La proposition et le développement de nouvelles représentations discriminantes d'images constituent une tâche difficile dans différents domaines liés à la vision par ordinateur et la reconnaissance des formes. Les méthodes traditionnellement utilisées pour la reconnaissance d'images reposent généralement sur une description statistique ou structurelle du contenu de ces dernières, sous la forme de caractéristiques visuelles telles que les contours, la couleur ou encore la texture. Une limite de ces approches vient du fait que ces différents types de caractéristiques (ainsi que leurs combinaisons potentielles) sont parfois trop peu discriminantes pour décrire efficacement des images composées d'objets complexes et/ou à fort contenu sémantique.

Ces dernières années, la disposition spatiale des objets (ou des différentes régions composant les objets) présents dans une image a reçu une attention particulière dans le domaine de l'analyse d'images. En effet, les relations spatiales entre les objets d'une scène imagée constituent une information particulièrement importante dans la perception humaine de la similarité entre des images. Par conséquent, ces relations spatiales peuvent être considérées comme des caractéristiques importantes pour reconnaître le contenu de l'image ou la nature de l'objet représenté par cette dernière. Cependant, à notre connaissance, elles restent assez peu utilisées dans la littérature, principalement parce qu'elles reposent souvent sur de fortes contraintes structurelles.

Parallèlement, les approches dites « sacs de caractéristiques » (*bags of features*) ont été proposées en vision par ordinateur pour exploiter efficacement les aspects discriminants des descripteurs locaux dans les images. De telles stratégies ont conduit à des résultats très prometteurs dans les tâches de classification d'images, mais l'un de leurs inconvénients majeurs réside souvent le manque d'intégration d'informations spatiales et structurelles, notamment parce que les images sont représentées comme des ensembles non ordonnés de descripteurs locaux.

Cet article, qui fait suite à des travaux récemment présentés dans [7], décrit une nouvelle approche de type sacs de caractéristiques, fondée sur des descripteurs de position relative, pour la reconnaissance d'images à fort contenu sémantique. La Section 2 présente des travaux connexes à cette étude. Notre première contribution (présentée en Sec-

tion 3) est un descripteur modélisant à la fois la forme et les relations spatiales entre les différents objets composant une image. Ce descripteur structurel est une extension du concept de Décomposition en Histogramme de Forces (FHD) [14, 6]. Notre deuxième contribution (présentée en Section 4) est un nouveau cadre d'apprentissage à partir du descripteur FHD, inspiré des stratégies de sacs de caractéristiques. L'originalité de cette approche consiste à construire un vocabulaire homogène de formes et de configurations spatiales liant plusieurs objets sémantiques à différentes échelles d'analyse. La Section 5 présente ensuite une étude expérimentale préliminaire, où un jeu de données d'images de scènes naturelles est considéré pour illustrer l'intérêt de l'approche proposée. Nous montrons notamment que la combinaison de cette méthode structurelle avec une approche plus classique à partir de descripteurs locaux permet de mieux reconnaître ces images de scènes à fort contenu sémantique. Enfin, une conclusion et des perspectives sont présentées dans la Section 6.

2 Travaux connexes

2.1 Relations spatiales

De nombreuses études ont été menées pour l'analyse des relations spatiales, dans un but commun de décrire la position relative d'objets représentés dans des images [2]. On peut distinguer, dans la littérature, deux grands axes de recherche reposant sur des concepts fortement duaux [20] : (i) l'évaluation de relations spatiales et (ii) la position relative d'un objet par rapport à un autre.

Dans le premier axe, une relation spatiale spécifique est considérée, par exemple « *à gauche de* », et une évaluation floue de cette relation est réalisée pour deux objets donnés. Par exemple, de nombreux travaux liés aux paysages flous [1] reposent sur ce type d'évaluation. Cette approche est basée sur une modélisation floue des relations spatiales directement dans l'espace image, en utilisant des opérations morphologiques. Les applications typiques incluent par exemple la reconnaissance de visages à partir de graphes [5], la segmentation d'images IRM du cerveau [9], ou encore la reconnaissance de textes manuscrits [12].

Dans le deuxième axe, la position relative d'un objet par rapport à un autre peut avoir sa propre représentation (d'où il est alors possible de dériver des évaluations de différentes relations spatiales). Un descripteur de position relative couramment utilisé est l'Histogramme de Forces [19] (F-Histogramme), qui est une généralisation de l'Histogramme d'Angles [21]. Les F-Histogrammes sont utilisés dans plusieurs domaines d'application tels que les descriptions linguistiques [18], la mise en correspondance de scènes [4] ou la recherche d'images par le contenu [24].

Dans le contexte particulier de la reconnaissance d'images représentant des objets complexes, les auteurs de [14, 6] ont proposé un descripteur structurel appelé Décomposition en Histogrammes de Forces (FHD). En partant des couches disjointes de pixels (obtenues à partir d'une segmentation) composant un objet de l'image, le but de ce descripteur est d'encoder les relations spatiales calculées entre tous les couples de couches structurelles, à partir d'un ensemble homogène de F-Histogrammes. Ces travaux ont montré l'intérêt de cette représentation basée sur des relations spatiales directionnelles pour décrire des objets structurés. Cependant, cette approche est limitée par différentes contraintes. Tout d'abord, il est nécessaire de fixer *a priori* le nombre de couches structurelles composant un objet présent dans une image, afin de pouvoir comparer des descripteurs de même tailles. Deuxièmement, la comparaison de descripteurs FHD est vue comme un problème coûteux d'appariement de graphes, impliquant des contraintes structurelles fortes sur les sous-parties des objets composant l'image et une importante sensibilité par rapport à l'étape de segmentation initiale.

2.2 Vers les sacs de relations spatiales

Depuis une dizaine d'années, les approches par sacs de caractéristiques (*bags of features*), ont fait l'objet de nombreuses attentions pour les tâches de reconnaissance d'objets et de classification d'images [25, 13]. Le modèle par sacs de caractéristiques typique utilise des descripteurs locaux (*e.g.*, SIFT [16] ou HOG [11]) calculés à partir d'une détection de points d'intérêt de l'image ou bien de manière dense sur l'ensemble de l'image, afin de construire un vocabulaire de « mots visuels », en appliquant un algorithme de *clustering*. Une image est alors représentée par un histogramme de composition de ces mots visuels, et ces vecteurs caractéristiques forment l'entité de base pour la classification d'images, en utilisant par exemple un algorithme d'apprentissage supervisé (*e.g.*, SVM [10] ou forêts aléatoires [3]). Un inconvénient inhérent aux approches par sacs de caractéristiques est le manque de prise en compte d'information spatiale dans la description des images. Cette limite provient essentiellement du fait que les images sont représentées comme des collections non ordonnées de descripteurs locaux. Bien que certaines tentatives aient émergé pour essayer d'intégrer des informations spatiales dans ces modèles [15], peu d'approches prennent en compte la structure globale et les positions relatives des objets formant le contenu de l'image.

Dans le domaine de la reconnaissance des symboles, des travaux récents [22, 23] ont introduit les sacs de relations (*bags of relations*), comme une manière originale de produire des vocabulaires de configurations spatiales. L'approche a été appliquée à un ensemble restreint de primitives visuelles spécifiques à ce domaine d'application (par exemple des cercles, coins ou extrémités). Dans ces travaux, notre objectif est de continuer ces efforts et d'exploiter ce type d'approches par sacs de relations, en les étendant à l'application plus générique de la reconnaissance d'images composées d'objets complexes et/ou à fort contenu sémantique.

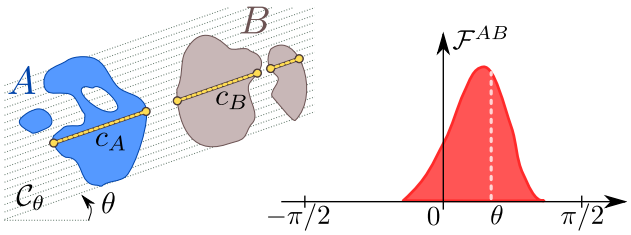

FIGURE 1 – Calcul d'un histogramme de force entre deux objets. L'histogramme $\mathcal{F}_{AB}$ décrit la position relative de l'objet A par rapport à l'objet B, en considérant toutes les directions.

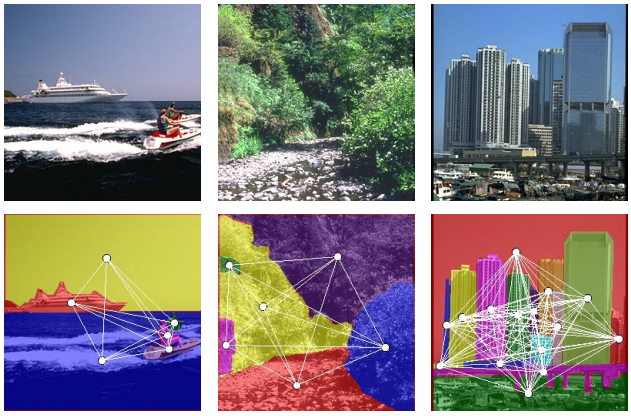

FIGURE 2 – Exemples illustratifs de graphes relationnels attribués (ARG) entre différents objets sémantiques composant des images de scènes naturelles.

3 Description structurelle d'images

Cette section présente l'approche proposée pour la description structurelle du contenu d'une image. Nous commençons ici par faire l'hypothèse que nous disposons d'une décomposition initiale de l'image en un ensemble d'objets sémantiques. L'obtention d'une telle décomposition n'est pas le problème traité dans le cadre de cet article mais peut être réalisée par un processus de partitionnement d'image (segmentation, classification) ou par une méthode de détection d'objets d'intérêt. À partir de cette décomposition structurelle de l'image, nous présentons une extension du descripteur de décomposition en histogrammes de forces (FHD) [14, 6] utilisé pour décrire le contenu de l'image, en caractérisant à la fois les formes et les relations spatiales entre les couples d'objets sémantiques composant le contenu de l'image.

3.1 Histogrammes de Forces

Un Histogramme de Forces (F-Histogramme) permet de décrire de manière directionnelle les relations spatiales entre un couple d'objets A et B représentés dans une image [19]. Sa construction repose sur la définition d'une force d'attraction entre les points de l'image. Étant donnés deux points situés à une distance d l'un de l'autre, leur force d'attraction est définie par $\varphi_r(d) = \frac{1}{d^r}$ où r caractérise le type de force traité. Lorsque $r = 0$, tous les points sont traités avec une importance égale (force constante), alors que lorsque $r = 2$, on donne alors plus d'importance aux points plus proches (force gravitationnelle). Ensuite, plutôt que d'étudier directement toutes les paires de points entre les deux objets, on considère la force d'attraction entre deux segments de droite. Soit I et J deux segments sur une droite orientée d'angle θ, D^θ_{IJ} la distance entre eux et $|.|$ la longueur d'un segment. La force d'attraction f_r de I par rapport à J est donnée par :

$$f_r(I,J) = \int_{D^\theta_{IJ}+|J|}^{|I|+D^\theta_{IJ}+|J|} \int_0^{|J|} \varphi_r(u-v)\,\mathrm{d}v\,\mathrm{d}u. \quad (1)$$

L'intersection d'une droite θ-orientée avec deux objets binaires A et B dans une image forme deux ensembles de segments appartenant à chaque objet : $\mathcal{C}_A = \cup\{I_i\}_{I=1..n}$ et $\mathcal{C}_B = \cup\{J_j\}_{j=1..m}$ (voir Figure 1). L'attraction mutuelle entre ces segments est définie comme suit :

$$F(\theta,\mathcal{C}_A,\mathcal{C}_B) = \sum_{I\in\mathcal{C}_A}\sum_{J\in\mathcal{C}_B} f_r(I,J). \quad (2)$$

Alors, l'ensemble de toutes les droites parallèles θ-orientées $\mathcal{C}_\theta$ traversant l'image, nous donne l'attraction globale $F^{AB}(\theta)$ entre A et B le long d'une direction θ. Enfin, le descripteur F-Histogramme $\mathcal{F}^{AB}$ est obtenu en calculant F^{AB} sur un ensemble d'angles $\theta \in [-\pi,+\pi]$, résumant ainsi la position relative de A par rapport à B. Le modèle des histogrammes de forces présente plusieurs propriétés d'invariance utiles et souvent recherchées dans un contexte de reconnaissance des formes : il sont invariants en translation et en homothétie, périodiques et quasi-invariants en rotation.

3.2 Décomposition en Histogrammes de Forces (FHD)

Étant donnée une décomposition initiale de l'image, l'idée clé du descripteur FHD est de caractériser les relations spatiales entre tous les couples d'objets composant le contenu des images en utilisant des F-Histogrammes. Lorsque le F-Histogramme est calculé entre une région de l'image et elle-même, il permet de décrire la forme de la région (F-Histogramme de « forme »), alors que pour deux régions différentes, le F-Histogramme décrit leur configuration spatiale (F-Histogramme de « relations spatiales »). L'image peut alors être représentée comme un graphe relationnel attribué (ARG) complet où les nœuds représentent les différentes régions de l'image (voir Figure 2). À chaque nœud est associé comme attribut le F-Histogramme décrivant sa forme, et à chaque arête le F-Histogrammes modélisant la position relative des deux nœuds reliés. Ainsi, une image composée de N objets (ou régions) est décrite par le graphe complet K_N composé de N nœuds dont les attributs sont des descripteurs de formes, et de $\frac{N(N-1)}{2}$ arêtes

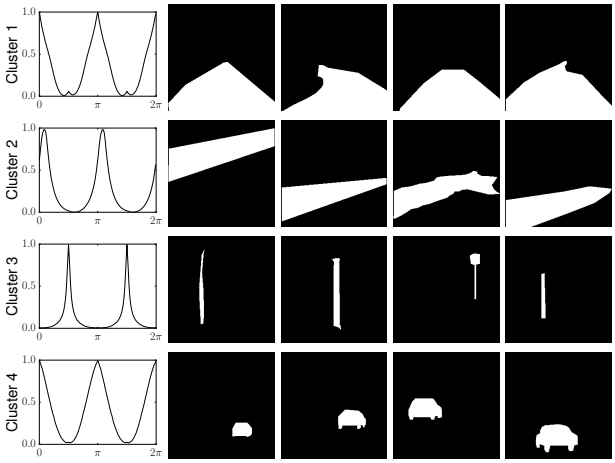

FIGURE 3 – Quelques échantillons représentatifs d'un vocabulaire de formes. Chaque ligne représente un mot du vocabulaire obtenu. (gauche) centroïde du *cluster* considéré ; (droite) objets représentatifs rattachés à ce centroïde.

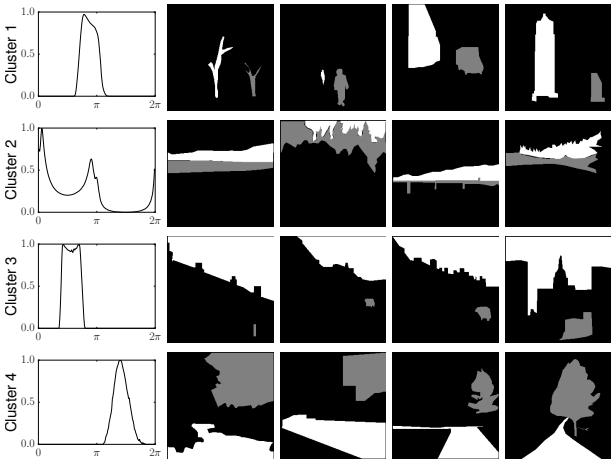

FIGURE 4 – Quelques échantillons représentatifs d'un vocabulaire de relations spatiales. Chaque ligne représente un mot du vocabulaire obtenu. (gauche) centroïde du *cluster* considéré ; (droite) configurations spatiales représentatives rattachées à ce centroïde. Les objets sont représentés en blanc et en gris.

dont les attributs sont des descripteurs de position relative.
Afin d'étendre le descripteur FHD à notre problématique de reconnaissance d'images de scènes naturelles, nous proposons de relâcher la contrainte structurelle de ce modèle ARG initial, c'est-à-dire de créer des ensembles sans ordre de F-Histogrammes entre tous les couples d'objets présents dans l'image. Ceci permet de surmonter les limitations du modèle reposant sur des graphes statiques énoncées précédemment en Section 2.1.
Ainsi, à partir de l'ARG de F-Histogrammes d'une image, nous obtenons deux ensembles différents de F-Histogrammes pour décrire l'image :

— l'ensemble S_{shapes} qui est composé des F-Histogrammes de « formes » (attributs des nœuds) ;
— l'ensemble $S_{\text{relations}}$ qui est composé des F-Histogrammes de « relations spatiales » (attributs des arêtes).

Les ensembles S_{shapes} et $S_{\text{relations}}$ calculés sur différentes images forment alors de façon homogène deux espaces caractéristiques indépendants dans lesquels la procédure d'apprentissage pour construire des sacs de caractéristiques sera réalisée.

4 Sacs de relations spatiales et de formes

Dans cette section, nous présentons notre cadre d'apprentissage basé sur les caractéristiques de relations spatiales et de formes précédemment décrites. Cette approche s'inspire des stratégies classiques par sacs de caractéristiques, qui sont généralement appliquées à partir de descripteurs locaux. Étant donné un ensemble d'images portant des contenus sémantiques différents, notre objectif est de construire un vocabulaire des formes, ainsi que des configurations spatiales les plus représentatives, apparaissant dans le contenu des images à différents niveaux d'analyse. Pendant la phase d'apprentissage, nous appliquons d'abord la stratégie de description structurelle d'images pour construire les ensembles S_{shapes} et $S_{\text{relations}}$ pour chaque image de l'ensemble d'apprentissage. Nous appliquons ensuite une stratégie de quantification vectorielle des espaces de caractéristiques induits par les descripteurs de relations spatiales et de formes, suivie d'une étape de regroupement pour finalement représenter chaque image par un histogramme de composition des mots du vocabulaire.
Pour la quantification vectorielle, nous utilisons l'algorithme de *clustering* K-Means [17] afin de construire des groupes de caractéristiques similaires dans l'ensemble d'apprentissage. Nous réalisons deux *clusterings* K-Means distincts : le premier avec l'ensemble accumulé S_{shapes} de F-Histogrammes de « formes », conduisant à des *clusters* de formes semblables au sein des images d'apprentissage (voir Figure 3) ; le second avec l'ensemble accumulé $S_{\text{relations}}$ de F-Histogrammes de « relations spatiales », construisant des *clusters* de configurations spatiales similaires entre objets sémantiques (voir Figure 4). D'une manière analogue aux mots visuels classiques (produits par le regroupement de descripteurs locaux), nous produisons un vocabulaire structurel de formes et de configurations spatiales qui apparaissent à travers les objets sémantiques contenus dans les images de l'ensemble d'apprentissage. Le nombre de clusters K_{shapes} et $K_{\text{relations}}$ déterminent les tailles respectives des vocabulaires de formes et de relations spatiales. L'influence de ces paramètres sera étudiée expérimentalement dans la Section suivante.
Une fois que les vocabulaires ont été construits, nous appliquons alors une étape de *pooling* : les images d'apprentissage sont ensuite représentées par des histogrammes

FIGURE 5 – Exemples d'images de scènes naturelles du jeu de données « *8 scènes* ». Chaque image est issue d'une classe différente.

de composition, modélisant le nombre d'occurrences de chaque mot du vocabulaire détecté dans les images (formes et relations spatiales). Pour une image de test donnée, nous appliquons alors la même stratégie : l'histogramme de composition est construit en attribuant ses F-Histogrammes aux mots de vocabulaire obtenus pendant la phase d'apprentissage. Finalement, la classification est effectuée à partir de machines à vecteurs de support (SVM) [10].
Étant donné que nous construisons deux vocabulaires distincts, nous produisons deux types d'histogrammes de composition : l'un pour les formes et l'autre pour les relations spatiales. On nomme *sacs de formes* (*BoS* pour *bags of shapes*) et *sacs de relations* (*BoR* pour *bags of relations*) la prise en compte respective des formes ou des relations spatiales seulement. Enfin, la concaténation des histogrammes de composition des formes et des relations spatiales s'appelle *sacs de formes et de relations* (*BoSR*).
Par ailleurs, même si la nature des informations modélisées par notre approche structurelle et par une approche classique de sacs de caractéristiques (à partir de descripteurs locaux) est différente, les procédures appliquées de *clustering* et de *pooling* aboutissent à des vecteurs de caractéristiques de même nature (histogrammes de composition), alors concaténables en vecteurs de caractéristiques hybrides. Cette stratégie de combinaison sera étudiée dans la section expérimentale de cet article.

5 Validations expérimentales

5.1 Jeu de données

Le jeu de données « *8 scènes* » est un ensemble d'images représentant des scènes naturelles en extérieur [1]. Il est composé d'un total de 2686 images de taille 256×256 réparties en 8 classes homogènes correspondant à différents types de scènes complexes (forêt, ville, montagne, etc.). Quelques exemples d'images pour chacune des classes sont présentées en Figure 5. Ce jeu de données est également accompagné d'annotations de référence des différents objets sémantiques (ciel, route, personne, bâtiment, etc.) qui composent les scènes. Ces annotations permettent de produire des masques binaires représentant précisément la localisation des objets, et qui peuvent être utilisés comme sous-parties structurelles des images. Ces annotations sont particulièrement utiles pour évaluer la performance de notre approche de description structurelle, car elles permettent de ne pas dépendre d'une étape de segmentation initiale qui peut s'avérer difficile. Par ailleurs, cela permet d'illustrer de manière concrète et intuitive la capacité de description de l'information spatiale, car en utilisant les masques des objets de cette manière, l'approche ne considère pas les valeurs colorimétriques des pixels de l'image, mais uniquement les configurations spatiales et les formes des objets la composant.

1. http ://people.csail.mit.edu/torralba/code/spatialenvelope/

5.2 Protocole expérimental

Les F-Histogrammes sont calculés sur un ensemble discret de 180 directions, avec une force constante, et son normalisés sur l'ensemble $[0, 1]$ pour obtenir une invariance à l'échelle des objets.
Nous appliquons alors l'approche par sacs de caractéristiques proposée pour classifier les images du jeu de données à partir des objets sémantiques représentés à l'intérieur de celles-ci. Les histogrammes de composition obtenus sont entraînés à partir de machines à vecteurs de support (SVM) [10] avec un noyau linéaire, en utilisant une approche *one-versus-all* pour la classification multi-classes. Nous appliquons une stratégie de validation croisée de type K-*Fold* avec $K = 10$. Ensuite, nous calculons plusieurs statistiques (taux de reconnaissance, précision et rappel), afin d'évaluer la qualité des résultats de la classification.
Par ailleurs, nous comparons nos résultats avec une approche de sacs de caractéristiques classique fondée sur l'utilisation de descripteurs locaux. Nous utilisons les caractéristiques HOG [11] de manière dense sur l'ensemble de l'image. Les caractéristiques HOG sont calculées sur 9 orientations discrètes et avec une fenêtre de taille 32×32 sans normalisation entre les fenêtres. La classification des images est effectuée en utilisant le même classifieur SVM linéaire que pour notre approche.

5.3 Résultats expérimentaux

La Figure 6 montre les taux de reconnaissance obtenus pour la classification des images du jeu de données, pour différentes tailles de vocabulaire, et pour les différentes approches proposées : BoS, BoR et BoSR. À partir de ces résultats, nous pouvons constater une meilleure performance pour l'approche BoR (relations spatiales) par rapport à l'approche BoS (formes uniquement). La combinaison de ces deux approches (BoSR), qui combine les deux types d'information, permet d'obtenir des taux de reconnaissance sensiblement supérieurs. Par ailleurs, nous observons une relative robustesse des taux de reconnaissance par rapport à l'évolution de la taille du vocabulaire. Ce résultat suggère que, pour ce jeu de données, la majorité de l'information est contenue dans un petit nombre de mots du vocabulaire.

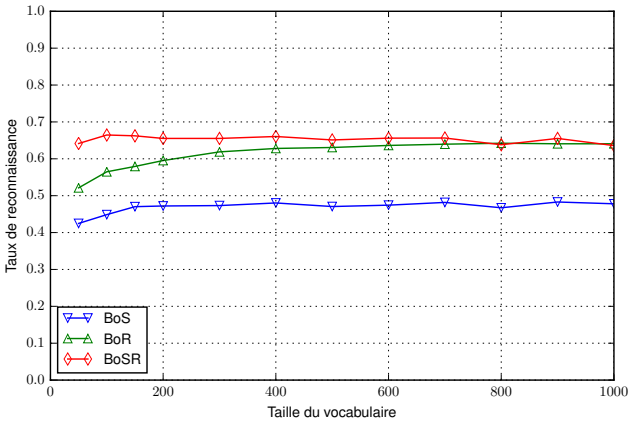

FIGURE 6 – Taux de reconnaissance obtenus pour les différentes approches proposées (BoS, BoR et BoSR), pour différentes tailles de vocabulaires allant de 50 à 1000.

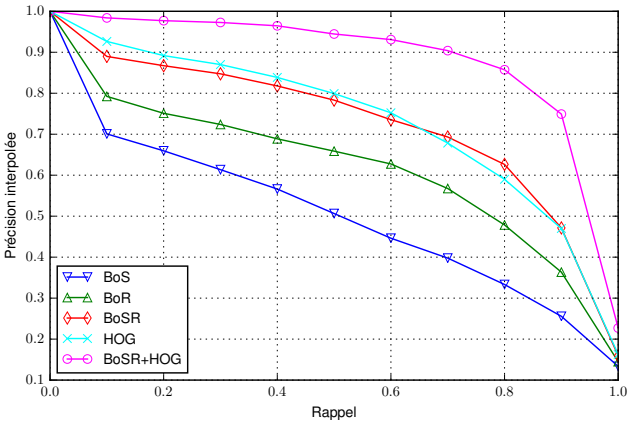

FIGURE 7 – Courbes de précision-rappel obtenues pour les différentes approches proposées (BoS, BoR et BoSR), pour l'approche comparative utilisant les caractéristiques HOG, ainsi que pour la combinaison des approches BoSR+HOG.

La Figure 7 présente les courbes de précision-rappel obtenues pour les différentes approches proposées (BoS, BoR et BoSR), ainsi que pour l'approche comparative utilisant les caractéristiques HOG. Nous présentons également les résultats obtenus pour la combinaison des approches BoSR et HOG. Les résultats sont obtenus pour une taille de vocabulaire fixée à 100 pour chacune des approches, dans un souci de compromis entre richesse de description et temps de calcul. Nous pouvons constater que l'approche BoSR obtient des résultats comparables à ceux obtenus par l'approche classique HOG. Enfin, nous observons une amélioration significative pour la combinaison BoSR+HOG. Ce résultat est particulièrement intéressant, car il suggère que des représentations hybrides entre informations locales et structurelles permettent de mieux reconnaître ce type d'images. Cette interprétation avait d'ailleurs été pressentie dans nos travaux précédents pour la classification d'images d'objets structurés [7]. Ainsi, ces validations confirment l'intérêt de combiner des stratégies d'apprentissage de descripteurs locaux et de descripteurs structurels pour améliorer les processus de reconnaissance.

Enfin, le Tableau 1 présente plus en détails les taux de reconnaissance obtenus pour les 8 différentes classes du jeu de données, en fonction des différentes approches considérées. À partir de ces résultats, on remarque notamment que la moitié des classes sont mieux reconnues par l'approche locale HOG, tandis que l'autre moitié des classes sont mieux reconnues par l'approche structurelle BoSR. Nous pouvons également de constater que, sauf pour une classe, la combinaison BoSR+HOG permet toujours d'améliorer le taux de reconnaissance. Ceci est particulièrement vrai pour la classe *opencountry*, qui est pourtant difficilement reconnue par les approches prises individuellement.

TABLEAU 1 – Comparaison des taux de reconnaissance obtenus pour les différentes classes du jeu de données.

	BoSR	HOG	BoSR+HOG
coast	53.89	**89.44**	80.56
forest	76.22	**90.24**	**92.38**
highway	**78.76**	44.79	**80.69**
insidecity	**87.66**	69.16	**92.53**
mountain	77.01	**86.90**	**90.11**
opencountry	**28.29**	12.68	**64.15**
street	**75.26**	44.67	**82.82**
tallbuilding	68.54	**71.63**	**85.11**

6 Conclusion

Nous avons présenté dans cet article une nouvelle approche, fondée sur des descripteurs de position relative, pour la reconnaissance d'images. La principale originalité de cette approche est la proposition d'une stratégie par sacs de caractéristiques permettant d'apprendre un vocabulaire homogène de formes et de configurations spatiales entre des objets sémantiques composant les images. Nous avons également montré que la combinaison de cette méthode structurelle avec un cadre classique de descripteurs locaux permet de mieux reconnaître des images de scènes complexes. Les résultats obtenus sur un jeu de données d'images de scènes naturelles mettent en évidence l'intérêt de telles représentations pour la classification d'images à fort contenu sémantique.

Par la suite, nous envisageons d'incorporer d'autres types de descripteurs de relations spatiales dans notre approche ; par exemple des relations topologiques, ou d'autres plus spécifiques comme l'entourement ou l'entrelacement d'objets [8], fournissant des éléments composites de vocabulaire dans cette stratégie d'apprentissage. Une autre perspective repose sur une méthode d'interprétation et d'adaptation des vocabulaires de configurations spatiales, dans le but d'effectuer des requêtes sémantiques de haut niveau sur la structure des images.

Références

[1] I. Bloch. Fuzzy relative position between objects in image processing : A morphological approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(7) :657–664, 1999.

[2] I. Bloch. Fuzzy spatial relationships for image processing and interpretation : a review. *Image and Vision Computing*, 23(2) :89–110, 2005.

[3] L. Breiman. Random Forests. *Machine Learning*, 45(1) :5–32, 2001.

[4] A. R. Buck, J. M. Keller, and M. Skubic. A memetic algorithm for matching spatial configurations with the histograms of forces. *IEEE Transactions on Evolutionary Computation*, 17(4) :588–604, 2013.

[5] R. M. Cesar, E. Bengoetxea, and I. Bloch. Inexact graph matching using stochastic optimization techniques for facial feature recognition. In *International Conference on Pattern Recognition (ICPR)*, volume 2, pages 465–468, 2002.

[6] M. Clément, M. Garnier, C. Kurtz, and L. Wendling. Color object recognition based on spatial relations between image layers. In *International Conference on Computer Vision Theory and Applications (VISAPP)*, pages 427–434, 2015.

[7] M. Clément, C. Kurtz, and L. Wendling. Bags of Spatial Relations and Shapes Features for Structural Object Description. In *International Conference on Pattern Recognition (ICPR)*, 2016.

[8] M. Clément, A. Poulenard, C. Kurtz, and L. Wendling. Directional Enlacement Histograms for the Description of Complex Spatial Configurations between Objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017 (in press).

[9] O. Colliot, O. Camara, and I. Bloch. Integration of fuzzy spatial relations in deformable models - Application to brain MRI segmentation. *Pattern Recognition*, 39(8) :1401–1414, 2006.

[10] C. Cortes and V. Vapnik. Support-Vector Networks. *Machine Learning*, 20(3) :273–297, 1995.

[11] N. Dalal and B. Triggs. Histograms of Oriented Gradients for Human Detection. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages 886–893, 2005.

[12] A. Delaye and E. Anquetil. Learning of fuzzy spatial relations between handwritten patterns. *International Journal on Data Mining, Modelling and Management*, 6(2) :127–147, 2014.

[13] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The pascal visual object classes (VOC) challenge. *International Journal of Computer Vision*, 88(2) :303–338, 2010.

[14] M. Garnier, T. Hurtut, and L. Wendling. Object description based on spatial relations between level-sets. In *International Conference on Digital Image Computing Techniques and Applications (DICTA)*, pages 1–7, 2012.

[15] S. Lazebnik, C. Schmid, and J. Ponce. Beyond Bags of Features : Spatial Pyramid Matching for Recognizing Natural Scene Categories. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages 2169–2178, 2006.

[16] D. G. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 60(2) :91–110, nov 2004.

[17] J. B. MacQueen. Some methods for classification and analysis of multivariate observations. In *Berkeley Symposium on Mathematical Statistics and Probability (BSMSP)*, pages 281–297, 1967.

[18] P. Matsakis, J. M. Keller, L. Wendling, J. Marjamaa, and O. Sjahputera. Linguistic description of relative positions in images. *IEEE Transactions on Systems, Man, and Cybernetics, Part B : Cybernetics*, 31(4) :573–88, 2001.

[19] P. Matsakis and L. Wendling. A new way to represent the relative position between areal objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(7) :634–643, 1999.

[20] P. Matsakis, L. Wendling, and J. Ni. A general approach to the fuzzy modeling of spatial relationships. In *Methods for Handling Imperfect Spatial Information*, pages 49–74. 2010.

[21] K. Miyajima and A. Ralescu. Spatial organization in 2D segmented images : Representation and recognition of primitive spatial relations. *Fuzzy Sets and Systems*, 65(2) :225–236, 1994.

[22] K. Santosh, B. Lamiroy, and L. Wendling. Integrating vocabulary clustering with spatial relations for symbol recognition. *International Journal on Document Analysis and Recognition*, 17(1) :61–78, 2014.

[23] K. Santosh, L. Wendling, and B. Lamiroy. BoR : Bag-of-Relations for Symbol Retrieval. *International Journal of Pattern Recognition and Artificial Intelligence*, 28(6), 2014.

[24] S. Tabbone and L. Wendling. Color and grey level object retrieval using a 3D representation of force histogram. *Image and Vision Computing*, 21(6) :483–495, 2003.

[25] J. Zhang, M. Marszalek, S. Lazebnik, and C. Schmid. Local features and kernels for classification of texture and object categories : A comprehensive study. *International Journal of Computer Vision*, 73(2) :213–238, 2007.