



HAL
open science

Bags of Spatial Relations and Shapes Features for Structural Object Description

Michaël Clément, Camille Kurtz, Laurent Wendling

► **To cite this version:**

Michaël Clément, Camille Kurtz, Laurent Wendling. Bags of Spatial Relations and Shapes Features for Structural Object Description. International Conference on Pattern Recognition (ICPR), Dec 2016, Cancún, Mexico. pp.1994-1999, 10.1109/ICPR.2016.7899929 . hal-01496896

HAL Id: hal-01496896

<https://hal.science/hal-01496896>

Submitted on 1 Jun 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Bags of Spatial Relations and Shapes Features for Structural Object Description

Michaël Clément, Camille Kurtz and Laurent Wendling
Université Paris Descartes, Sorbonne Paris Cité, LIPADE (EA 2517)
45 rue des Saints-Pères, 75006 Paris, France
michael.clement@parisdescartes.fr

Abstract—We introduce a novel bags-of-features framework based on relative position descriptors, modeling both spatial relations and shape information between the pairwise structural subparts of objects. First, we propose a hierarchical approach for the decomposition of complex objects into structural subparts, as well as their description using the concept of Force Histogram Decomposition (FHD). Then, an original learning methodology is presented, in order to produce discriminative hierarchical spatial features for object classification tasks. The cornerstone is to build an homogeneous vocabulary of shapes and spatial configurations occurring across the objects at different scales of decomposition. An advantage of this learning procedure is its compatibility with traditional bags-of-features frameworks, allowing for hybrid representations of both structural and local features. Classification results obtained on two datasets of images highlight the interest of this approach based on hierarchical spatial relations descriptors to recognize structured objects.

I. INTRODUCTION

The development of discriminative image representations is a challenging problem in various fields related to image analysis and pattern recognition. Classical methods dedicated to the recognition of complex objects usually rely on a structural or statistical description of the object content, summarizing different image features such as outer contour, geometry or texture and color effects. A limit of these methods is that these different types of features (and combinations) are sometimes not discriminant enough to successfully describe image contents composed of complex objects.

In recent years, the spatial disposition of objects in a scene (or between object sub-components) has received much attention in the domain of image analysis. In fact, it can be stated that structural relations between image components are fundamental in the human perception of image similarity. Therefore, the spatial relations between the regions composing an object can be considered as important features to recognize the nature of the object itself. However, as far as we know they are seldom used for object recognition, mostly because they often suffer from strong structural constraint issues.

In parallel, bags-of-features strategies have been proposed in the computer vision community to efficiently exploit the discriminative aspects of local features in images. Such strategies have led to encouraging results in image classification tasks, but one of their inherent downside is the lack of spatial information, because images are represented as orderless collections of local features.

In this article, we present a novel bags-of-features framework based on spatial relations descriptors for the description and recognition of complex objects. Section II reviews some related work in the context of this paper. Our first contribution, presented in Section III, is a hierarchical approach for the decomposition of complex objects, as well as their multilevel structural description using the concept of Force Histogram Decomposition (FHD) [14], [7]. Our second contribution, presented in Section IV, is the introduction of a novel learning framework, inspired from bags-of-features strategies, to learn discriminative hierarchical spatial features for object classification tasks. The main originality of this approach is to build an homogeneous vocabulary of shapes and spatial configurations occurring in the objects at different hierarchical scales. In Section V, we propose an experimental study, where two datasets of complex objects are considered to illustrate the robustness of our proposed approach. We show that the combination of this structural method with a classical local features framework allows to better recognize complex objects. Finally, a conclusion that emphasizes the perspectives of this work is found in Section VI.

II. RELATED WORK

A. Spatial Relations

Many studies have been conducted for the analysis of spatial relations in different application domains of pattern recognition and computer vision, with the common objective of describing the relative position of objects in images [3]. We can distinguish in the literature two main research axes based on strong dual concepts [19]: the concept of spatial relation, and the relative position of an object with regards to another.

In the first axis, a spatial relation such as “*to the left of*” is considered, and a fuzzy evaluation of this relation is obtained for given two objects. For instance, the fuzzy landscape framework [2] is oriented towards this type of evaluations. This approach is based on a fuzzy modeling of spatial relations directly in the image space, using morphological operations. Typical applications include for example graph-based face recognition [6], brain segmentation from MRI [8], or handwritten text recognition [12].

In the second axis, the relative position of an object with regards to another one can have a representation of its own, from which it is then possible to derive evaluations of spatial relations. A typical relative position descriptor is

the Force Histograms [18], which was itself inspired by Angle Histograms [20]. Force Histograms are used in several application domains such as linguistic descriptions [17], scene matching [5] or content-based image retrieval [23], [14], [7].

In this particular context of image retrieval, the authors of [14], [7] proposed a structural object descriptor called Force Histogram Decomposition (FHD). The key idea of this descriptor is to encode the pairwise spatial relations between disjoint layers of pixels composing an object, obtained with a segmentation strategy, using an homogeneous set of F-Histograms. These works showed the interest of this representation based on directional spatial relations for structured object description. However, this approach suffers from different problems. First, it is required to *a priori* fix the number of object subparts to run the decomposition step of this descriptor. Secondly, the comparison of two objects described with FHD descriptors is viewed as a (costly) graph matching problem – where nodes represent object subparts and edges represent pairwise spatial relations – implying strong structural constraints on the object subparts and a high sensitivity to segmentation issues.

B. Towards Bags-of-Relations

Bags-of-features (BoF) strategies, have recently attracted numerous research attentions for object recognition and image classification tasks [26], [13]. The typical BoF model uses local features (*e.g.*, SIFT [15] or HOG [11] descriptors), either based on sparse interest points or in a dense grid, to build a vocabulary of visual words, with a clustering technique. An image is then represented by a composition histogram of such visual words, and these feature vectors form the basic visual entity for image classification, using a supervised machine learning procedure (*e.g.*, SVM [10] or Random Forests [4]). An inherent downside of the BoF frameworks is the lack of spatial information, because images are represented as orderless collections of local features. Some approaches have tried to incorporate spatial information, but only few approaches use the inner structure of objects.

In the field of symbol recognition, recent works [21], [22] introduced bags-of-relations (BoR), an original way to produce composite vocabularies of spatial configurations. The approach was applied on a well-controlled set of visual primitives specific to the application domain (*e.g.*, circles, corners or extremities). In this work, our goal is to leverage this type of approach, extending it to the more generic application of complex objects in color images.

III. HIERARCHICAL DESCRIPTION OF OBJECTS

This section presents our framework for the description of complex objects depicted in images. First, we introduce how to decompose an object into its inner structural subparts with a hierarchical image segmentation. Then, given this structural decomposition, we present an original extension of the Force Histogram Decomposition (FHD) descriptor [14], [7] used to characterize both the shapes and the pairwise spatial relations between the hierarchical subparts of the object.

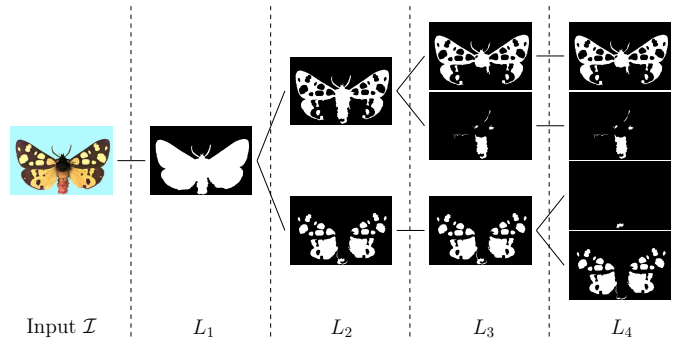


Fig. 1. Hierarchical decomposition of a typical structured object into its inner subparts. For illustration purposes, only the top four levels of the hierarchy are represented.

A. Hierarchical Object Segmentation

Given an object to be described, our objective is first to decompose it into its different structural subparts according to the image content. We define a structural subpart of an object as a set of pixels in the image, potentially composed of multiple connected components (*e.g.*, two wheels form a single structural subpart of a motorbike). As object subparts can emerge at various levels of image interpretation, we propose to consider a hierarchical image segmentation strategy to extract the structural subparts at different scales. The hierarchical object segmentation algorithm starts with an initial over-segmentation of the image and uses this level as a base for the construction of subsequent significant levels.

From the input image, we first apply the Mean Shift segmentation algorithm [9], which yields a set of s non-overlapping segments S_i (modeling a partition of the input image), delineating small regions exposing homogeneous color properties. The Mean Shift algorithm works with a set of intuitive parameters allowing to control the granularity of the segmentation. We then consider that each segment S_i represents an initial structural subpart of the object.

To reconstruct the structural layers of the object at different complementary scales, we then apply the Ascendant Hierarchical Clustering (AHC) algorithm [24] (using the Euclidean metric, with an average linkage criteria) on the RGB values of the s segments obtained with the Mean Shift segmentation. One advantage of considering a clustering algorithm, as opposed to a more classical hierarchical segmentation algorithm with spatial constraints [1], is that it enables to build structural layers potentially composed of multiple connected components. The AHC procedure builds a hierarchy of image segments (by successive binary merging of the most similar, but not necessarily adjacent, subparts of the object), that can be represented as a binary tree structure, commonly called a dendrogram \mathcal{D} of s levels. Each level of \mathcal{D} models a particular scale of structural layers of the object: the “leaf” level corresponds to the s initial structural subparts while the root level is the cluster that contains one single structural subpart composed of all the image segments. Each intermediate level

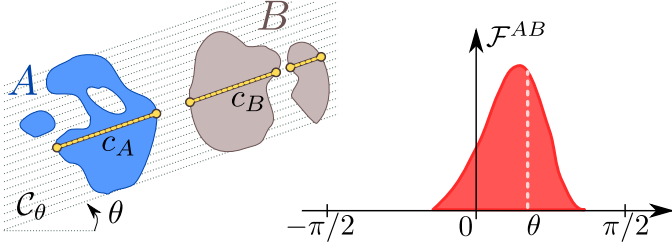


Fig. 2. Computation of a Force Histogram between two objects. The histogram \mathcal{F}_{AB} describes the relative position of the object A with regards to the object B , considering all directions.

L_i with $i \in [1, s]$ of \mathcal{D} can be represented as an independent set of structural subparts ($|L_i| = i$) of the object whose union is a partition of the input image.

These dendrogram levels model a hierarchical decomposition of the object into different (but complementary) structural inner subparts. Such a hierarchical decomposition of a typical structured object is presented in Fig. 1, showing the top four levels of decomposition. The next step will consist in homogeneously characterizing both the shape and the pairwise spatial relations between these subparts to provide a multilevel description of the object.

B. Force Histogram Decomposition (FHD)

A Force Histogram (thereafter noted F-Histogram) allows to describe the directional spatial relations between a couple of objects A and B depicted in an image [18]. It relies on the definition of a force of attraction between points. Given two points located at a distance d from each other, their force of attraction is defined by $\varphi_r(d) = \frac{1}{d^r}$ where r characterizes the kind of force processed. When $r = 0$, all points are treated with equal importance (constant force), whereas when $r = 2$, more importance is given to closer points (gravitational force). Instead of directly studying all pairs of points between the two objects, the force of attraction between two segments is considered. Let I and J be two segments on a line of angle θ , D_{IJ}^θ the distance between them and $|\cdot|$ the segment length. The force of attraction f_r of I with regard to J is given by:

$$f_r(I, J) = \int_{D_{IJ}^\theta + |J|}^{|I| + D_{IJ}^\theta + |J|} \int_0^{|J|} \varphi_r(u - v) dv du. \quad (1)$$

Given two binary objects A and B , a θ -oriented line in the image forms two sets of segments belonging to each object: $\mathcal{C}_A = \cup\{I_i\}_{i=1..n}$ and $\mathcal{C}_B = \cup\{J_j\}_{j=1..m}$ (see Fig. 2). The mutual attraction between these segments is defined as:

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

Then, the set of all θ -oriented parallel lines \mathcal{C}_θ going through the whole image, gives us the global attraction $F^{AB}(\theta)$ between A and B along a direction θ .

Finally, the F-Histogram descriptor \mathcal{F}^{AB} is obtained by computing F^{AB} onto a set of angles $\theta \in [-\pi, +\pi]$, summarizing the relative position of A with regards to B . The

model highlights useful theoretical properties: it is invariant with regards to translations and scaling transformations, quasi-invariant to rotations, and periodic. Besides, a notable property is that F-Histograms are bilinear, that is, for any objects A_1, A_2, B_1 and B_2 , we have:

$$\mathcal{F}^{(A_1 \cup A_2)(B_1 \cup B_2)} = \mathcal{F}^{A_1 B_1} + \mathcal{F}^{A_1 B_2} + \mathcal{F}^{A_2 B_1} + \mathcal{F}^{A_2 B_2}. \quad (3)$$

This property is particularly interesting when considering objects composed of multiple connected components, allowing for a particularly efficient computation of pairwise F-Histogram in our hierarchical agglomerative framework.

Given a “flat” (non-hierarchical) decomposition of an object in an image into a fixed number of subparts, the key idea of the initial FHD descriptor was to encode the pairwise spatial relations between all these image layers using F-Histograms. When applied to each layer with itself, the spatial self-relations encode shape information (“shape” F-Histograms) whereas for two different structural subparts, the spatial relations encode relative structural aspects (“spatial relations” F-Histograms). The object can finally be represented as a complete attributed relational graph (ARG) of F-Histograms modeling the spatial relations (and self-relations) between all the possible couples of image layers. To extend the FHD descriptor to our hierarchical object decomposition, we propose to relax the structural constraint of this initial ARG model, that is, we create orderless sets of F-Histograms between all the couples of object structural subparts at each scale of image decomposition. This allows both to benefit from the hierarchical segmentation as we obtain a coarse-to-fine description of the object, and to overcome the limitations of the static graph model.

To this end, starting from a given stage L_k of the image dendrogram (where $k \geq 2$ denotes the considered maximum depth of decomposition), we compute two different sets of F-Histograms to describe the object. On the one hand, the set S_{shapes} is composed of shape F-Histograms, each being computed between a subpart and itself, resulting for each object in a set of shape descriptors at different levels of decomposition L_i with $i \in [1, k]$. This results in a total of $|S_{\text{shapes}}| = \frac{k(k+1)}{2}$ F-Histograms (*i.e.*, i shape descriptors for each stage L_i). On the other hand, the set $S_{\text{relations}}$ is composed of spatial relation F-Histograms, each being computed between all pairs of subparts at all levels of decomposition. This results in a total of $|S_{\text{relations}}| = \sum_{i=2}^k \frac{i(i-1)}{2}$ F-Histograms (*i.e.*, $\frac{i(i-1)}{2}$ spatial relations descriptors for each stage L_i). Note that, given the bilinear property of F-Histogram, the generation of these two sets is particularly efficient, because we only need to compute F-Histograms at the maximum depth stage L_k , and then the F-Histograms for the successive stages are automatically obtained from the previous ones (without running the algorithm again).

The two sets S_{shapes} and $S_{\text{relations}}$ of F-Histograms, computed on different objects, homogeneously represent two independent feature spaces in which the learning procedure to build bags of features will be realized.

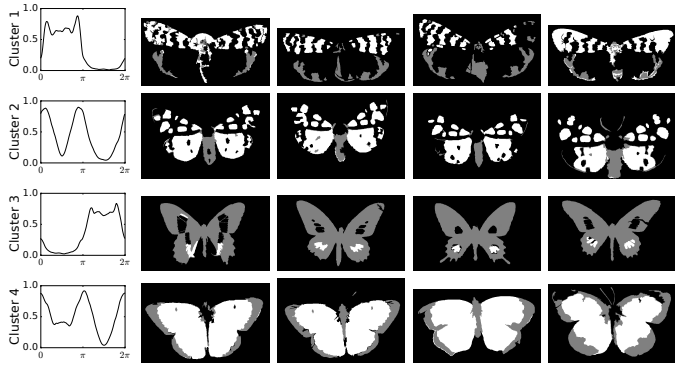


Fig. 3. Some representative samples from our vocabulary of spatial configurations issued from the vector quantization step. Each row represents a cluster, or vocabulary word, obtained using our bags-of-features learning strategy: (left) centroid of the cluster, (right) representative object subparts attached to this centroid. The two object subparts whose spatial relations are computed are depicted in white and gray.

IV. BAGS OF SPATIAL RELATIONS AND SHAPES

In this section, we introduce our learning framework based on the previously described hierarchical spatial relations and shapes features. This approach is inspired by traditional bags-of-features strategies usually applied with local descriptors. Given a training set of images containing objects from different categories, our objective is to build a vocabulary of typical shapes, as well as spatial configurations occurring across the objects subparts at different scales. During the learning phase, we first apply the hierarchical object description strategy to construct sets S_{shapes} and $S_{\text{relations}}$ for each object of the training set. We then apply a vector quantization strategy in the feature space spanned by our spatial relations and shapes features, followed by a pooling step to finally represent each object by an histogram of vocabulary words composition.

For vector quantization, we use the K-Means clustering algorithm [16] in order to build clusters of similar features in the training set. We perform two distinct K-Means clusterings: the first one with the accumulated set S_{shapes} of shape F-Histograms, therefore building clusters of similar occurring subparts in the dendrograms of the training images (see Fig. 4), while the second one is with the accumulated set $S_{\text{relations}}$ of spatial relations F-Histograms, building clusters of similar spatial configurations between subparts (see Fig. 3), at different scales of the dendrograms. In a similar fashion to the “visual words” produced by the clustering of local features, we produce a structural vocabulary of shapes and spatial configurations occurring across the objects in the training set. The number of clusters K_{shapes} and $K_{\text{relations}}$ determine the respective sizes of the shapes and spatial relations vocabularies. The influence of these parameters will be studied in the following experimental section.

Once the vocabularies have been built, we apply a pooling step: training images are subsequently represented by sparse composition histograms, modeling the count of each detected vocabulary word in the objects (shapes and spatial relations).

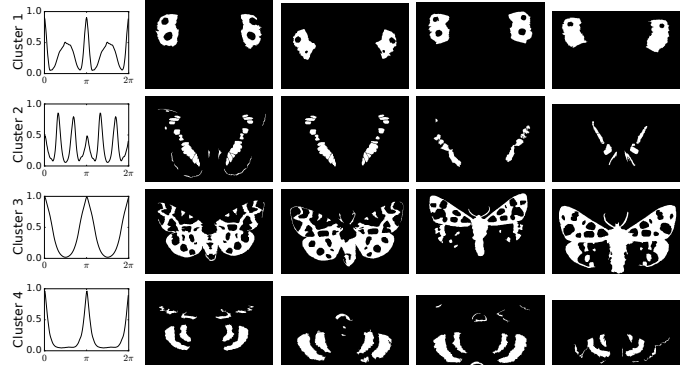


Fig. 4. Some representative samples from our vocabulary of shapes issued from the vector quantization step. Each row represents a cluster, or vocabulary word, obtained using our bags-of-features learning strategy: (left) centroid of the cluster; (right) representative object subparts attached to this centroid.

A given test image follows the same hierarchical decomposition strategy, and its composition histogram is built by assigning its F-Histograms to the vocabulary words obtained during the training phase. Finally, classification is performed using Support Vector Machines (SVM). Given the fact that we build two distinct vocabularies, we also produce two types of composition histograms: one for shapes and one for spatial relations. We name *bags-of-shapes* (BoS) and *bags-of-relations* (BoR) the respective consideration of shapes or spatial relations only. Subsequently, The concatenation of shapes and spatial relations composition histograms is called *bags-of-shapes-and-relations* (BoSR).

Besides, even though the semantics of our structural approach and of local features are different (descriptors are extracted from images with very different methodologies), the applied clustering and pooling procedures result in feature vectors of the same nature (sparse composition histograms), allowing them to be combined into hybrid feature vectors by concatenating them. This combination strategy is studied in the experimental section.

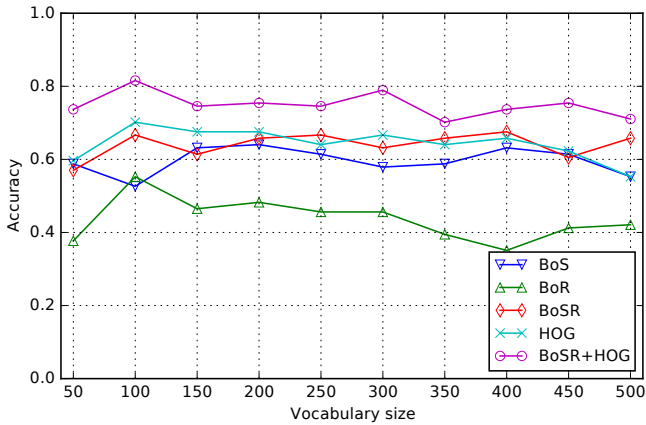
V. EXPERIMENTAL RESULTS

A. Datasets

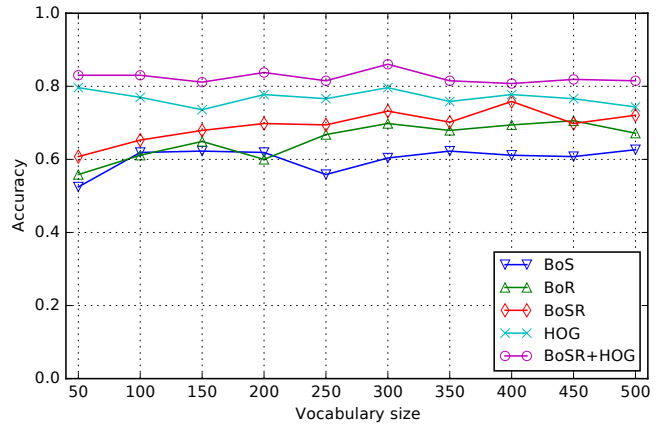
We used two datasets of color images: a collection of butterfly images and a collection of motorbike images. Samples of these datasets are shown in Fig. 6. As we focus on the proposition of a new descriptor for object classification, we validate our approach on specific datasets where objects are depicted on an homogeneous background, avoiding object detection issues. The recognition of complex scenes in natural color images will constitute a perspective of this work.

The Peale dataset is a subset of the Peale collection¹. Our representative subset is composed of 114 color butterfly images grouped into 11 classes according to their species, each class containing a variable number of butterflies (5 to 20 samples per class). Butterflies are a typical case where inner

¹<http://clade.ansp.org/entomology>



(a) Peale dataset



(b) Motorbikes dataset

Fig. 5. Classification accuracy scores obtained for different vocabulary sizes on the two datasets. Vocabulary size correspond to the number of clusters formed during the training phase. BoS: Bags-of-Shapes, BoR: Bags-of-Relations, BoSR: Bags-of-Shapes-and-Relations, HOG: Histogram of Oriented Gradients.



Fig. 6. Samples from the Peale and the Motorbikes datasets: butterfly from different species and motorbikes from different categories.

spatial relations are a distinguishing feature making the wings patterns a direct link with the species.

The Motorbikes dataset is a subset of a collection of various motorcycles from the side made available by Caltech². Our subset is composed of 265 motorbikes images with an homogeneous background. As for the butterflies, motorbikes are an interesting case where the shape and spatial disposition of the wheels, body and engine parts are discriminative features. We manually labeled the motorbikes into five categories: cross, cruiser, moped, sport and standard.

B. Experimental Protocol

The parameters of the Mean Shift segmentation have been fixed according to a previous study with these datasets in [7]. The spatial and range radiuses are set to $(h_s, h_r) = (8, 8)$ and the minimum size of the regions is set to $M = 100$. Then, our hierarchical FHD descriptors are computed starting with a maximum depth of $k = 5$, resulting in a structural coarse-to-fine description of the objects (from 1 to 5 subparts). F-Histograms along a discrete set of 180 directions, with a constant force, and are normalized onto $[0, 1]$ to achieve scale invariance across the hierarchy of subparts.

²<http://vision.caltech.edu/archive.html>

For each dataset, we apply our proposed bags-of-features approach to perform object classification. We train the composition histograms using Support Vector Machines (SVM) [10] with a linear kernel, using a *one-against-all* approach for multiclass classification. Cross validation is achieved following a *leave-one-out* strategy. We then compute several statistics (accuracy, precision, recall and F1-score), in order to evaluate the quality of the classification results.

As for the comparative study, we compare our results with the following approaches:

- The initial FHD descriptors for color images [7]. We report the best results obtained for this approach (fixed decomposition of objects into 3 layers) ;
- The GFD shape descriptors [25], which only encode global shape information (with $R = 4$ radial and $T = 9$ angular frequencies). For this approach, classification is performed using k -NN and the Euclidean metric ;
- Local HOG [11] features sampled in a dense grid, and pooled with a bags-of-features framework. We used HOG with 9 discrete orientations, and cell size of 32×32 without cross-cell normalization. Classification is performed using the same linear SVM classifier as our approach.

C. Quantitative Results

Fig. 5 shows the classification accuracy scores obtained on the two datasets, for varying vocabulary sizes on our proposed approaches: BoS, BoR, and BoSR. We also present the results obtained for HOG, as well as the combination BoSR+HOG. From these results, we can observe a relative robustness of the accuracy scores with regards to the size of the vocabulary. Besides, an important result is that the combination of shapes and spatial relations (BoSR) seems to always improve the performance as opposed to shapes (BoS) and spatial relations (BoR) considered individually. The BoSR method yields comparable results to the semantically different HOG features. In addition, another important result

TABLE I
COMPARATIVE CLASSIFICATION RESULTS ON THE PEALE DATASET.

	FHD [7]	GFD [25]	HOG [11]	BoSR	BoSR+HOG
F1-score	0.77	0.44	0.70	0.70	0.78
Accuracy	0.77	0.44	0.70	0.71	0.80

is that the BoSR+HOG combination significantly increases the classification accuracy. This shows that, on these datasets, the combination of semantically different features (structural spatial relations and local feature descriptors) allows to improve the recognition process.

For comparative purposes, Tab. I and II show results for different approaches (GFD and FHD) on the two datasets. For HOG, BoSR and BoSR+HOG, we report the best results obtained in the previous vocabulary size study. The results show the interest of considering spatial relations features (FHD and BoSR) as opposed to global shape descriptors (GFD) to describe structural objects. They also show that our proposed BoSR+HOG hybrid representation outperforms the results of FHD on these datasets.

VI. CONCLUSION

In this article, we proposed a new object description approach based on hierarchical spatial relations features. The main originality of this proposition is a bags-of-features strategy allowing to learn a vocabulary of shapes and spatial configurations between structural subparts composing the objects. We also proposed hybrid representations, combining classical local features with our structural spatial relations descriptors. The results obtained on two datasets highlight the interest of such representations for the classification of complex objects depicted in images.

In the future, we plan to incorporate other types of spatial relations descriptors into our approach (*e.g.*, topological or other specific relations such as the surrounding or the interlacement of objects), providing composite elements of vocabulary in this learning strategy. Another perspective relies on an adaptive pruning of the spatial configurations vocabulary, with the goal to perform high-level semantic queries from images.

REFERENCES

- [1] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik. Contour detection and hierarchical image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(5):898–916, 2011.
- [2] 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.
- [3] I. Bloch. Fuzzy spatial relationships for image processing and interpretation: A review. *Image and Vision Computing*, 23(2):89–110, 2005.
- [4] L. Breiman. Random Forests. *Machine learning*, 45(1):5–32, 2001.
- [5] 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.
- [6] R. M. Cesar, E. Bengoetxea, and I. Bloch. Inexact graph matching using stochastic optimization techniques for facial feature recognition. In *Proceedings of the International Conference on Pattern Recognition – ICPR 2002*, volume 2, pages 465–468, 2002.

TABLE II
COMPARATIVE CLASSIFICATION RESULTS ON THE MOTORBIKES DATASET.

	FHD [7]	GFD [25]	HOG [11]	BoSR	BoSR+HOG
F1-score	0.71	0.66	0.79	0.71	0.82
Accuracy	0.71	0.66	0.80	0.72	0.83

- [7] M. Clément, M. Garnier, C. Kurtz, and L. Wendling. Color object recognition based on spatial relations between image layers. In *Proceedings of the International Conference on Computer Vision Theory and Applications – VISAPP 2015*, pages 427–434, 2015.
- [8] 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.
- [9] D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5):603–619, 2002.
- [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 2005*, volume 1, pages 886–893, 2005.
- [12] A. Delays 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 *Proceedings of the IEEE International Conference on Digital Image Computing Techniques and Applications – DICTA 2012*, pages 1–7, 2012.
- [15] D. G. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 60(2):91–110, nov 2004.
- [16] J. B. MacQueen. Some methods for classification and analysis of multivariate observations. In *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability – BSMSP 1967*, pages 281–297, 1967.
- [17] 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.
- [18] 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.
- [19] 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.
- [20] 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.
- [21] 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.
- [22] 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.
- [23] 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.
- [24] J. H. Ward. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58(301):236–244, 1963.
- [25] D. Zhang and G. Lu. Shape-based image retrieval using generic Fourier descriptor. *Signal Processing: Image Communication*, 17(10):825–848, 2002.
- [26] 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.