

# Color Object Recognition based on Spatial Relations between Image Layers

Michaël Clément, Mickaël Garnier, Camille Kurtz and Laurent Wendling

*Université Paris Descartes, LIPADE – SIP (EA 2517), 45 rue des Saints-Pères, 75006 Paris, France*

**Keywords:** Object Recognition, Spatial Relations, Force Histograms, Mean Shift Segmentation, Shape Matching.

**Abstract:** The recognition of complex objects from color images is a challenging task, which is considered as a key-step in image analysis. Classical methods usually rely on structural or statistical descriptions of the object content, summarizing different image features such as outer contour, inner structure, or texture and color effects. Recently, a descriptor relying on the spatial relations between regions structuring the objects has been proposed for gray-level images. It integrates in a single homogeneous representation both shape information and relative spatial information about image layers. In this paper, we introduce an extension of this descriptor for color images. Our first contribution is to consider a segmentation algorithm coupled to a clustering strategy to extract the potentially disconnected color layers from the images. Our second contribution relies on the proposition of new strategies for the comparison of these descriptors, based on structural layers alignments and shape matching. This extension enables to recognize structured objects extracted from color images. Results obtained on two datasets of color images suggest that our method is efficient to recognize complex objects where the spatial organization is a discriminative feature.

## 1 INTRODUCTION

Object recognition from color images is an important topic in image analysis that has been largely studied in various imaging domains. The object recognition task is generally performed in three steps (Andreopoulos and Tsotsos, 2013): (1) extraction of salient regions using a segmentation algorithm; (2) description of these regions with quantitative features modeling different types of information (*e.g.*, color, texture or shape); (3) recognition of the objects of interest by classifying the characterized regions. A limit of this approach is that these different types of features (and their combinations) are sometimes not discriminant enough to successfully describe the content of complex objects, in particular when they are represented by color images.

For the last decades, the spatial disposition of objects in a scene (or between object subparts) has received much attention in the domain of image analysis (Santosh et al., 2012; Morales-González and García-Reyes, 2013). Moreover, it has been 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 im-

portant features to recognize the nature of the object itself. However, as far as we know they are rarely used for color object recognition.

Literature about spatial relations for image analysis can be structured in two main categories: qualitative and quantitative approaches. Qualitative approaches use symbolic relations such as positioning (left, right, below, etc.) or topological relations (inside, outside, etc.) (Egenhofer, 1989; Inglada and Michel, 2009). When dealing with complex objects, it is necessary to capture a precise description of their different structuring patterns, and to characterize both large-scale and low-scale directional relations. Besides, depending on the content meaning, these patterns may be represented by unconnected subsets of pixels. Therefore, in our context the spatial relations cannot be summarized in a symbolic manner. Quantitative approaches gather methods that precisely describe the relative positions between objects. Fuzzy quantitative methods are popular in different application domains such as spatial reasoning in medical images (Bloch and Ralescu, 2003; Bloch, 2005) or handwritten symbol recognition (Delays and Anquetil, 2011). These methods produce a fuzzy landscape per considered potential direction, but combining them in order to capture the omnidirectional spa-

tial organization of possibly sparse objects is not obvious. Another quantitative strategy is based on the model of *force histograms* (Matsakis and Wendling, 1999), thereafter noted F-Histograms. This model handles sparse object descriptions and summarizes their relative position in a single histogram. Basically, an F-Histogram between two objects is a circular distribution measuring their relative attraction along every desired directions.

Based on this model, the authors of (Garnier et al., 2012) proposed a new object descriptor called F-Histogram Decomposition (FHD). The key idea of this descriptor is to encode the pairwise spatial relations between disjoint layers of pixels composing an object using F-Histograms. This work led to encouraging results, showing the interest of considering an homogeneous representation of directional spatial relations for object description. However, this approach has been proposed only for gray-level images, and the decomposition strategy does not consider the content of the image to create the layers, but “hard” pixel layers based on pre-defined image intensity values. Consequently, this approach cannot straightforwardly deal with complex color images.

In this article, we present an extension of this approach for the recognition of complex objects in color images. The contributions of this article are twofold. First, we propose an extension of the FHD descriptor for color textured images by considering a robust object decomposition strategy. This “content-based” decomposition strategy relies on a combination of the Mean Shift segmentation algorithm coupled to the K-Means clustering algorithm enabling to extract binary image layers supposed to represent relevant object subparts. Second, we present different strategies for the comparison and matching of FHD descriptors. Given two objects to be compared, the idea is to align the different image layers based on various types of information (*i.e.*, luminance, shape), before comparing their corresponding F-Histograms. These matchings, which involve F-Histogram matrix reorganizations, enable to enhance FHD comparisons supported by dissimilarity computations in the context of color images. These contributions allow for robust structured object recognition from complex images.

This article is organized as follows. Sec. 2 introduces our method for color object description based on spatial relations between image layers. Sec. 3 describes the experimental validations carried out with this method on two color image datasets. Conclusions and perspectives will be found in Sec. 4.

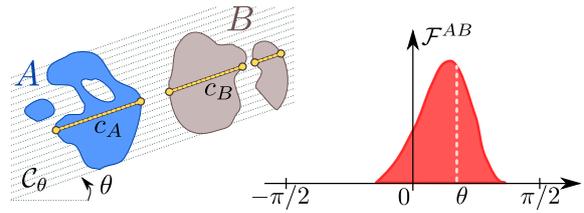


Figure 1: The force of attraction between  $A$  and  $B$  along the direction  $\theta$  is the integral sum of forces computed on longitudinal cuts ( $C_A, C_B$ ).

## 2 METHODOLOGY

First, we briefly recall the F-Histograms model (Sec. 2.1). Afterwards, we present our strategy for the decomposition of color objects into structural layers (Sec. 2.2). Then, the FHD object descriptor for color images is presented (Sec. 2.3). Finally, we propose different matching strategies for the efficient comparison of such FHD descriptors (Sec. 2.4).

### 2.1 Force Histograms Background

An F-Histogram allows to assess the directional spatial relations between two binary objects (Matsakis and Wendling, 1999). 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  $\theta$ ,  $D_{IJ}^\theta$  the distance between them and  $|\cdot|$  the segment length. The force of attraction  $f_r$  of  $I$  with regard to  $J$  is given by:

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

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

$$F(\theta, C_A, C_B) = \sum_{I \in C_A} \sum_{J \in C_B} f_r(I, J). \quad (2)$$

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

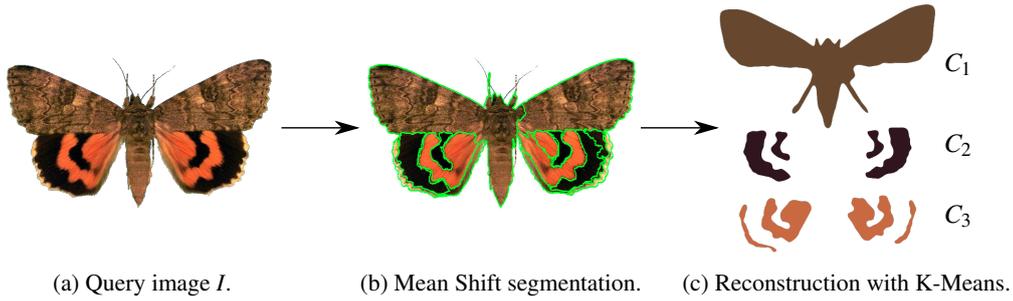


Figure 2: Illustration of the color image decomposition into structural layers.

Finally, the F-Histogram  $\mathcal{F}^{AB}$  is obtained by computing  $F^{AB}$  onto a set of angles  $\theta \in [-\pi, +\pi]$ , summarizing the relative position of  $A$  and  $B$ . By axiomatic definitions of the function  $F$ , the following properties, useful to characterize similar patterns, can be verified (Matsakis and Wendling, 1999): translation invariance as objects are processed independently of their location in the image, symmetry considering opposite directions, scale invariance if the histograms are normalized and rotation invariance (after circular shifts), because the approach is isotropic.

## 2.2 Decomposition into Image Layers

In order to capture the inner spatial relations structuring an object, we first have to decompose it into distinct subparts. To this end, we use a decomposition strategy based on a segmentation algorithm. However, in our context, the subparts of the object can be composed of disconnected sets of pixels. Consequently, we couple our segmentation strategy to a clustering algorithm enabling to reconstruct the different structural image layers of the object.

### 2.2.1 Color Image Segmentation

Color image segmentation is still an open research issue and no generic robust algorithm exists yet. In this work, we choose to employ the Mean Shift algorithm (Comaniciu and Meer, 2002) which has been successfully used during the last decade for both the classification and the segmentation of textured color images. The Mean Shift segmentation algorithm performs as follows. Let  $I$  be an input color image (see Fig. 2 (a)). For a given pixel of  $I$ , this algorithm builds a set of neighboring pixels within a given spatial radius  $h_s$  and color range  $h_r$ . The spatial and color center of this set is then computed and the algorithm iterates with this new spatial and color center. A third parameter  $M$  allows to define the minimum size of the regions to be kept after segmentation. The Mean Shift segmentation algorithm produces a set of  $R$  non-overlapping

segments  $S_i$  modeling a partition of  $I$ , and exposing homogeneous color properties (see Fig. 2 (b)).

### 2.2.2 Color Image Layers Reconstruction

Resulting from the Mean Shift segmentation, a structural subpart of the object can be disconnected in different segments (*e.g.*, see the orange parts on the butterfly wings in Fig. 2 (b)). Thus, our goal is to reconnect the segments sharing similar color intensity values into the same structural layers. Consequently, we apply the classical K-Means clustering algorithm on the  $R$  image segments  $S_i$  issued from the Mean Shift segmentation, to automatically reconstruct the structural layers of the object. The K-Means algorithm has been used with  $N$  seeds to produce  $N$  clusters of segments sharing similar color properties in the RGB space. These clusters of segments model the  $N$  structural layers  $C_i$  of the object (see Fig. 2 (c)). Note that any clustering algorithm could be used.

## 2.3 F-Histogram Decomposition

Once the object has been decomposed into  $N$  structural layers, we then compute an F-Histogram between every pair of layers ( $C_i, C_j$ ) according to the strategy proposed in (Garnier et al., 2012). The color F-Histogram Decomposition (FHD) of an object depicted in an image  $I$  is thus the set defined as:

$$\underbrace{\{\mathcal{F}_r^{C_i C_i}\}_{\forall i \in \{1..N\}}}_{\text{shape}} \cup \underbrace{\{\mathcal{F}_r^{C_i C_j}\}_{\forall (i,j) \in \{1..N\}^2, j > i}}_{\text{spatial relations}}. \quad (3)$$

These F-Histograms can be represented as a matrix (see Fig. 3) encoding  $N$  *shape* descriptors (when  $i = j$ ), and  $N(N-1)/2$  *spatial relations* descriptors between pairwise layers (when  $i > j$ ). An F-Histogram computed on a layer with itself induces a natural overlapping, only allowing to use a constant force of attraction (*i.e.*,  $r = 0$ ). In this study, for comparability and normalization purposes, we also use a

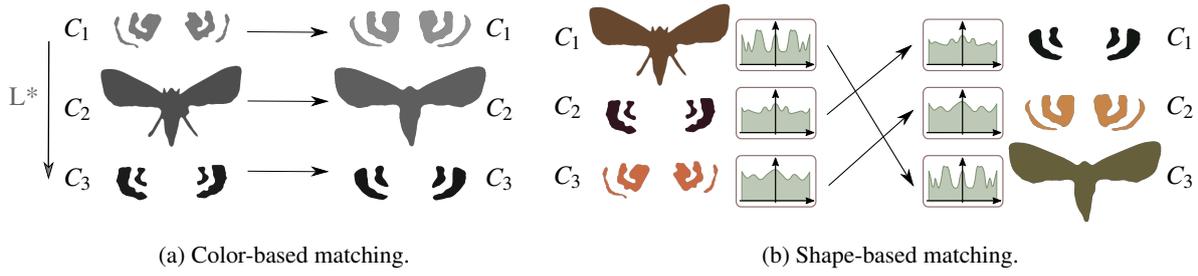
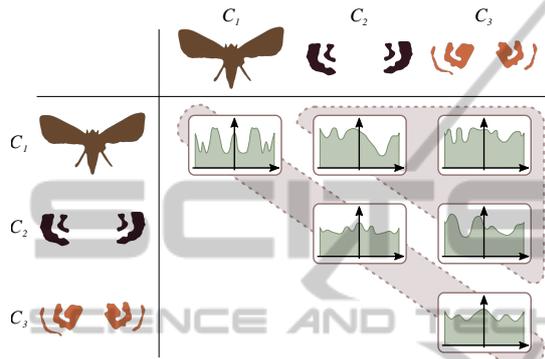


Figure 4: Illustration of the two proposed FHD matching strategies.


 Figure 3: Force Histogram Decomposition (FHD) of an object. An F-Histogram is computed between every pair of layers ( $C_i, C_j$ ) issued from the image decomposition step.

constant force for spatial relations F-Histograms, resulting in an homogeneous descriptor.

As a consequence of F-Histograms properties, FHD descriptors are naturally translation invariant and symmetric. Depending on the application requirements, scale invariance is obtained by normalizing the F-Histograms, and rotation invariance can be pursued by estimating their mode or by minimizing the distance between shifted FHD descriptors.

## 2.4 Comparison of FHD Descriptors

Comparing FHD descriptors can be done by evaluating the dissimilarity between the two matrices of F-Histograms, using a distance metric. Based on the results of (Garnier et al., 2012), we use the  $\chi^2$  distance to compare pairwise F-Histograms. Given two F-Histograms  $\mathcal{F}_A$  and  $\mathcal{F}_B$  computed along  $\theta_{max}$  directions, the  $\chi^2$  distance between them is defined as:

$$d_{\chi^2}(\mathcal{F}_A, \mathcal{F}_B) = \sum_{i=0}^{\theta_{max}} \frac{(\mathcal{F}_A(i) - \mathcal{F}_B(i))^2}{\mathcal{F}_A(i) + \mathcal{F}_B(i)}. \quad (4)$$

As noted before, the FHD descriptor includes both shape descriptions (on the diagonal) and pairwise spatial relations (on the upper triangle). The semantic meaning of these descriptions being different, we can

split the overall distance metric into two distinct parts, that is, for two FHD descriptors  $Q$  and  $T$ :

$$\mathcal{D}(Q, T) = \alpha \mathcal{D}_{shape}(Q, T) + (1 - \alpha) \mathcal{D}_{spatial}(Q, T), \quad (5)$$

where the distance between shapes and the distance between spatial relations are respectively defined as:

$$\mathcal{D}_{shape}(Q, T) = \frac{1}{N} \sum_{i=1}^N d_{\chi^2}(\mathcal{F}_Q^{C_i C_i}, \mathcal{F}_T^{C_i C_i}), \quad (6)$$

$$\mathcal{D}_{spatial}(Q, T) = \frac{2}{N(N-1)} \sum_{i=1}^N \sum_{j=i+1}^N d_{\chi^2}(\mathcal{F}_Q^{C_i C_j}, \mathcal{F}_T^{C_i C_j}), \quad (7)$$

and with  $\alpha \in [0, 1]$  balancing the weight given to shape descriptions compared to spatial relations. In the experimental section of this paper, we study the influence of this parameter on the retrieval results.

This approach for comparing FHD descriptors is only relevant if the underlying image layers structuring the two objects are correctly aligned in the two matrices. On the contrary, if the underlying image layers are misaligned (see Fig. 5), the induced FHD descriptors will lead to meaningless pairwise comparisons. This case is particularly likely to occur since we are dealing with color images, given that there is no natural order between image layers associated with intensity values in the RGB space.

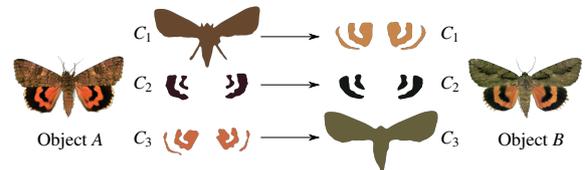


Figure 5: Illustration of a typical case where similar objects described with FHD descriptors can be incorrectly compared due to a misalignment of their structural layers.

Therefore, before applying the pairwise dissimilarity measure described above, it seems crucial to employ a matching strategy between the structural

layers of the objects. Finding such a matching between two FHD descriptors is a combinatorial problem (for  $N$  layers, there exists  $N!$  possible matchings), but the complexity of it can be effectively reduced by making assumptions on the underlying image layers. Thereafter, we propose different matching strategies based on specific alignments of the image layers. In our approach, FHD descriptors are computed only once, and the matching step only requires reorganizations of the F-Histograms matrices.

**Color-based Matching.** This matching strategy is based on the assumption that corresponding layers in the two objects will be similar in terms of colorimetry. To this effect, the image layers of each FHD descriptor are sorted according to the luminance value of the pixels composing the clusters issued from the decomposition step. An illustration of the color-based matching strategy is shown in Fig. 4 (a). This strategy being based on the decomposition step only, it can be applied prior to the computation of the descriptors. Therefore, no dynamic reorganization is required and the corresponding F-Histograms in the two descriptors can be compared in a simple pairwise manner.

**Shape-based Matching.** Another approach when comparing two FHD descriptors is to assume that the corresponding layers in the two objects will be similar in terms of shape. Fig. 4 (b) illustrates the principle of this matching strategy: each layer is matched with the one presenting the closest shape F-Histogram. To apply this strategy, we can assimilate it to an optimization problem where the goal is to minimize the overall distance  $\mathcal{D}_{shape}(Q, T)$  between shape F-Histograms in two FHD descriptors  $Q$  and  $T$ . The exact solution requires the enumeration of the  $N!$  matching possibilities and can be considered as an optimal matching (from an optimization point of view). It is obviously not suitable for FHD descriptors of arbitrary number of layers. An approximate solution consists in matching the layers sequentially in a greedy manner. That is, for each layer of  $Q$ , we pick the closest layer of  $T$  in terms of distance between their shape F-Histograms. From an algorithmic point of view, this only requires a quadratic computational time, and can be considered as a good compromise. These shape-based matching strategies (optimal and greedy) require a dynamic reorganization of the F-Histogram matrices during the comparison process.

The choice of a matching strategy is strongly application dependent, and each strategy presents its downsides and benefits. While the color-based strategy is simple and can be considered as a pre-



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

processing step, it will fail if the color is not a discriminative feature (*e.g.*, if two objects have a similar structure but different colors). On the other hand, the idea of matching the layers according to their shape is more relevant, but with a higher computational complexity and an increased sensibility to noise.

### 3 EXPERIMENTS

We involved the presented framework in object recognition and retrieval tasks from color images.

#### 3.1 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 recognition, we choose to validate this approach on specific datasets where the objects are depicted on an homogeneous background, avoiding any object detection issues.

The Peale dataset is a subset of the Peale collection<sup>1</sup> and is composed of 318 color butterfly images grouped in 28 classes according to their species, each class containing a variable number of butterflies. Butterflies are a typical case where inner spatial relations are a distinguishing feature making the wings patterns a direct link with the species. The same database limited to grayscale images was previously used in (Garnier et al., 2012) to introduce FHD descriptors and to assess the descriptive strength of the spatial relations.

The Motorbikes dataset is a subset of a collection of various motorcycles from the side made available by Caltech<sup>2</sup>. 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

<sup>1</sup><http://clade.ansp.org/entomology>

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

Table 1: Evolution of the recognition rate  $\mathcal{R}$  for different decomposition methods, by varying the number of layers  $N$ .

Decomposition / $N$	$\mathcal{R}$ (Peale)					$\mathcal{R}$ (Motorbikes)				
	2	3	4	5	6	2	3	4	5	6
Grayscale level-sets	26.1	30.7	31.9	37.4	40.8	65.2	59.1	58.1	55.6	63.1
Mean Shift ( $h_r = 4$ )	42.8	45.0	44.0	42.8	45.0	69.8	<b>70.9</b>	63.8	63.8	63.0
Mean Shift ( $h_r = 8$ )	<b>47.5</b>	35.8	42.1	38.4	35.5	67.2	64.9	63.0	54.0	54.3
Mean Shift ( $h_r = 12$ )	39.9	32.4	36.2	31.1	25.2	65.3	63.4	67.2	67.2	66.0

Table 2: Evolution of the recognition rate  $\mathcal{R}$  for the different matching strategies, by varying the number of layers  $N$ .

Matching strategy / $N$	$\mathcal{R}$ (Peale)					$\mathcal{R}$ (Motorbikes)				
	2	3	4	5	6	2	3	4	5	6
Color-based	42.8	45.0	44.0	42.8	45.0	69.8	70.9	63.8	63.8	63.0
Greedy shape-based	41.2	45.0	34.3	34.3	30.8	67.9	64.9	62.6	54.0	54.3
Optimal shape-based	42.1	40.6	39.3	35.2	33.0	69.8	66.8	64.2	67.2	66.0

manually labeled the motorbikes into five categories: cross, cruiser, moped, sport and standard.

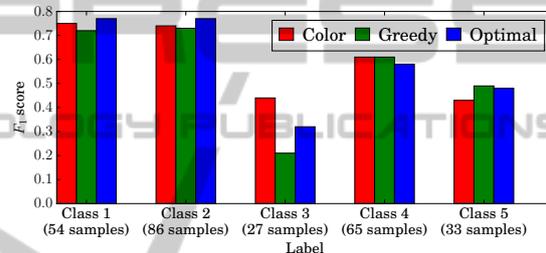
### 3.2 Protocol and Setup

For each dataset, object recognition is performed using the nearest neighbor strategy. The label of the closest image in the feature space spanned by FHDs is assigned to the object to be recognized. We used a leave-one-out cross validation strategy, that is, for each test image, the remaining of the dataset serves as training data. The quality of the results is then evaluated using classical quality indexes: the global recognition rate  $\mathcal{R}$ , and the class-by-class F-Measure  $F_1$ .

The robustness of our approach with regard to the decomposition step has been evaluated by studying the influence of the Mean Shift parameters ( $h_s$ ,  $h_r$  and  $M$ ), as well as the number of clusters  $N$  used for K-Means, producing different scales and levels of object decompositions, and resulting in FHD descriptors of different sizes. The minimum size of the regions has been set to  $M = 100$ , and the spatial bandwidth has also been set to  $h_s = 8$  (neighborhood of  $17 \times 17$  pixels, see (Comaniciu and Meer, 2002)), according to the average size of the images in the datasets. For global evaluation, we assigned an equal weight ( $\alpha = 0.5$  in Eq. 5) to shapes and spatial relations. The influence of this parameter is also studied in a class-by-class manner. Besides, all F-Histograms are computed along 180 directions, regularly spanning the  $[0, 2\pi[$  interval.

### 3.3 Results

**Influence of the Decomposition Step.** Tab. 1 presents the recognition rates obtained on the two

Figure 7: Class-by-class  $F_1$  scores obtained for the three matching strategies on the Motorbikes dataset.

datasets for different object decomposition methods. The first row presents the results obtained by considering the original version of FHDs applied on gray-level images using quantized level-sets (Garnier et al., 2012). The remaining rows show the results obtained by considering our extension of FHDs for color images. We observe that our extension performs better than the original gray-levels version. This confirms our hypothesis that considering a content-based decomposition strategy yields more relevant image layers. Besides, according to the different Mean Shift results, we can deduce that for these two specific datasets, it is more suitable to adopt an oversegmentation strategy.

**Influence of the Matching Strategy.** Tab. 2 presents the recognition rates for the different matching strategies on the two datasets. The first row shows the results obtained with the color-based matching strategy while the remaining rows present the results obtained with the shape-based matching strategies (greedy and optimal). On the Peale dataset, we can observe that the color-based strategy performs globally better than the shape-based ones for all values of

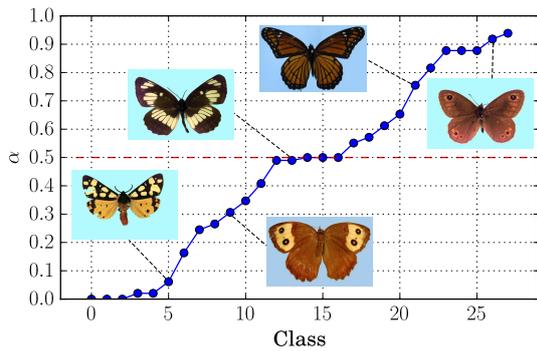


Figure 8: Best  $\alpha$  weights obtained for each class of the Peale dataset (with  $N = 3$ ). Classes are sorted by increasing values of  $\alpha$ . Lower values indicate a stronger weight on spatial relations, while higher values indicate a stronger weight on shape descriptions.

$N$ . This may be due to the fact that for most species of butterflies, the color of each layer is one of the most discriminative feature. On the contrary, on the Motorbikes dataset the best scores are mostly obtained with the optimal shape-based strategy. These results confirm that the shape-based strategy is more adapted to recognize objects where color is less discriminative. On this dataset, the greedy version of the shape-based matching produces comparable results to the optimal one, but with a much lower computational cost. Fig. 7 also present the class-by-class  $F_1$  scores obtained for the different matchings on the Motorbikes dataset, showing notably that the greedy strategy does not always perform worse than the optimal one.

**Influence of the  $\alpha$  Parameter.** We also studied the impact of the parameter  $\alpha$  representing the weight given to shapes compared to spatial relations (see Eq. 5). Instead of setting  $\alpha = 0.5$  as in the previous experiments, we evaluated the best local weights for each class of the dataset. We exhaustively computed the class-by-class recognition results for varying values of  $\alpha \in [0, 1]$  with a step of 0.02. For each class  $i$ , we record the  $\alpha_i$  values leading to the best local recognition rate for this specific class. Fig. 8 presents the result of this study on the Peale dataset. This figure highlights the idea that depending on the class, spatial relations or shape descriptions can be more or less discriminant features. Among the 28 classes of this dataset, about one third is more oriented towards spatial information, another third towards shape information, and the rest is balanced between both.

**Qualitative Evaluation.** Fig. 9 shows some representative object retrieval results. Color-based matching was used on the Peale dataset (with  $N = 3$ ), while



Figure 9: Retrieval results for the Peale and the Motorbikes datasets. Query images are depicted in the left column. Disimilarity rankings go from lowest (left) to highest (right).

shape-based matching was used on the Motorbikes dataset (with  $N = 4$ ).

**Comparative Study.** We compared the results obtained in this study to the ones obtained with the GFD (Zhang and Lu, 2002) and dSIFT (Bosch et al., 2006) descriptors with the same experimental protocol. The GFD is derived by applying two-dimensional Fourier transform on a polar-raster sampled image. Note that GFDs exclusively encode shape information. The computation of GFDs requires to set the radial and the angular resolutions ( $R$  and  $T$ ). In our experiments, GFDs are computed with  $R = 4$  and  $T = 9$ , thus giving a signature of 36 bins. Dense SIFTs are extracted with a step of 16 pixels and at two scales, 4 and 8, giving both local information and a more global one. For each keypoint of the query image, the matching algorithm searches for the best keypoint in the target image. If the resulting match gives a good contrast, this point vote goes to the target image. The finally matched image is the one with the highest votes. Tab. 3 presents the recognition rates obtained

Table 3: Recognition rates  $\mathcal{R}$  for dSIFT, GFD and FHD.

	GFD	dSIFT	FHD
<b>Peale</b>	28.6	43.4	<b>47.5</b>
<b>Motorbikes</b>	65.7	50.9	<b>70.9</b>

with these two methods, compared to the best results obtained with FHD descriptors. The results obtained with both GFDs and dSIFTs are slightly lower than the ones obtained with the proposed color FHD descriptors, showing that FHDs are more suited to recognize complex structured objects where spatial organization is a discriminative feature.

## 4 DISCUSSION

This article has introduced a novel approach for object recognition in color images. The main originality of this approach lies in the extension of an object descriptor using both shape and relative spatial information about the subparts of the object to be recognized. These subparts are obtained by considering a content-based decomposition relying on the combination of an image segmentation algorithm and a clustering algorithm. This decomposition strategy makes it possible to deal with the problem raised by the textural properties of the objects in color images. Experimental validations on two datasets have shown that describing objects with such shape and spatial relations features between structural image layers yields interesting recognition results.

Our work presents some limitations. From a methodological point of view, the image decomposition approach could be affected by the clustering step, which is based on color information. For instance, if different subparts of an object share the same color, it will lead to inconsistent image decompositions. In addition, the number of resulting clusters has to be set *a priori* and can be not adapted to the described object. From an applicative point of view, the datasets used were relatively small and present high intra-class variance making it difficult to accurately evaluate the behavior of our method.

In the future, we plan to enhance the FHD descriptor by considering a multi-layer strategy where the number of layers extracted per image may be automatically adapted depending on the image content. We also plan to involve this framework into biomedical object recognition tasks where the object of interest can be mixed up with the background.

## ACKNOWLEDGEMENTS

This work was sponsored by the French National Research Agency project SPIRIT #11-JCJC-008-01.

## REFERENCES

- Andreopoulos, A. and Tsotsos, J. K. (2013). 50 Years of object recognition: Directions forward. *Comput. Vis. Image Und.*, 117(8):827–891.
- Bloch, I. (2005). Fuzzy spatial relationships for image processing and interpretation: A review. *Image Vision Comput.*, 23(2):89–110.
- Bloch, I. and Ralescu, A. L. (2003). Directional relative position between objects in image processing: A comparison between fuzzy approaches. *Pattern Recogn.*, 36(7):1563–1582.
- Bosch, A., Zisserman, A., and Muoz, X. (2006). Scene classification via pLSA. In *Proc. of ECCV 2006*, volume 3954 of *LNCS*, pages 517–530.
- Comaniciu, D. and Meer, P. (2002). Mean shift: A robust approach toward feature space analysis. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(5):603–619.
- Delays, A. and Anquetil, E. (2011). Fuzzy relative positioning templates for symbol recognition. In *Proc. of IEEE ICDAR 2011*, pages 1220–1224.
- Egenhofer, M. J. (1989). A formal definition of binary topological relationships. In *Foundations of Data Organization and Algorithms*, volume 367 of *LNCS*, pages 457–472.
- Garnier, M., Hurtut, T., and Wendling, L. (2012). Object description based on spatial relations between level-sets. In *Proc. of IEEE DICTA 2012*, pages 1–7.
- Inglada, J. and Michel, J. (2009). Qualitative spatial reasoning for high-resolution remote sensing image analysis. *IEEE Trans. Geosci. Remote Sens.*, 47(2):599–612.
- Matsakis, P. and Wendling, L. (1999). A new way to represent the relative position between areal objects. *IEEE Trans. Pattern Anal. Mach. Intell.*, 21(7):634–643.
- Morales-González, A. and García-Reyes, E. (2013). Simple object recognition based on spatial relations and visual features represented using irregular pyramids. *Multimed. Tools Appl.*, 63(3):875–897.
- Santosh, K., Lamiroy, B., and Wendling, L. (2012). Symbol recognition using spatial relations. *Pattern Recogn. Lett.*, 33(3):331–341.
- Zhang, D. and Lu, G. (2002). Shape-based image retrieval using Generic Fourier Descriptor. *Signal Process. Image*, 17(10):825–848.