

Description d'objets en couleurs à partir des relations spatiales entre régions structurelles

Michaël Clément, Camille Kurtz, Laurent Wendling

Université Paris Descartes, LIPADE – SIP (EA 2517)
45 rue des Saints-Pères, 75006 Paris, France
prenom.nom@parisdescartes.fr

Résumé. Les méthodes classiques de reconnaissance d'objets à partir d'images reposent généralement sur une description statistique ou structurelle du contenu des objets, sous la forme de caractéristiques visuelles telles les contours, la couleur ou encore la texture. De telles caractéristiques sont alors considérées pour des tâches de classification ou de recherche d'images similaires. Dans cet article, nous présentons un nouveau descripteur d'objets complexes représentés dans des images en couleurs. L'originalité de ce descripteur réside dans une représentation homogène d'attributs de formes et de relations spatiales entre régions structurelles extraites des objets à partir d'une méthode robuste de décomposition. Nous proposons également différentes stratégies de comparaison de ces descripteurs, basées sur une mise en correspondance des régions structurelles des objets. Les résultats obtenus sur plusieurs bases d'images en couleurs montrent que les relations spatiales entre les différentes régions composant des objets complexes constituent des caractéristiques intéressantes pour leur description.

1 Introduction

1.1 Contexte

La reconnaissance d'objets complexes à partir d'images en couleurs est une tâche difficile qui est considérée comme une étape clé dans le domaine de l'analyse et de la fouille d'images. Les méthodes de reconnaissance d'objets reposent généralement sur une description automatique des objets représentés dans les images. Dans le cas où l'on s'intéresse à des objets en particulier présent dans les images (et non pas à l'ensemble de leur contenu), les tâches de reconnaissance sont généralement organisées en trois étapes (Andreopoulos et Tsotsos, 2013) : (1) détection et extraction de régions d'intérêt dans l'image en fonction de différents critères, à l'aide d'un algorithme de segmentation ; (2) description de ces régions à l'aide de caractéristiques quantitatives modélisant différents types d'information (forme, couleurs, texture, etc.) (Ilea et Whelan, 2011) ; (3) reconnaissance des objets d'intérêt par classification ou *clustering* des régions décrites.

Une des principales limites de ces approches dites *basées régions* vient du fait que ces différents types de caractéristiques visuelles conduisent souvent à des ensembles d'indices relativement hétérogènes qui peuvent devenir difficiles à combiner de manière efficace (Jolion,

Description d'objets à partir des relations spatiales entre régions

2001). De plus, ces indices ne permettent généralement pas de décrire correctement des objets complexes et structurés, d'autant plus lorsque ceux-ci sont représentés dans des images couleurs.

Parallèlement, l'organisation spatiale des différents objets dans une scène, ou des différentes régions d'intérêt d'un même objet, constitue une information particulièrement importante dans la perception humaine de la similarité entre les images. Par conséquent, dans notre contexte, les relations spatiales entre les régions composant des objets peuvent être considérées comme des caractéristiques discriminantes pour reconnaître la nature de ces objets. Cependant, ces caractéristiques spatiales sont rarement utilisées pour la reconnaissance d'objets à partir d'images en couleurs.

1.2 État de l'art et proposition

Dans la littérature, les méthodes décrivant le positionnement relatif d'objets dans des images peuvent être réparties en deux grandes familles d'approches : les approches qualitatives et les approches quantitatives.

Les méthodes de description qualitative des relations spatiales s'inscrivent souvent dans un contexte plus général de « *raisonnement qualitatif* » plutôt issu du domaine de l'intelligence artificielle. Les travaux de ce type s'intéressent notamment au raisonnement spatio-temporel du point de vue de la cognition et de la perception humaine (Freeman, 1975; Egenhofer, 1989; Cohn et al., 1997), et les applications principales de ces modèles se situent souvent dans le domaine de la télédétection (Inglada et Michel, 2009; Belarte et al., 2013). Cependant, ces méthodes offrent des représentations de l'information spatiale qui sont difficilement intégrables dans des processus de reconnaissance de formes. De plus, elles ont souvent recours à des simplifications grossières de la forme des objets (par exemple, réduction des objets à leur barycentre ou leur rectangle englobant), ou bien se ramènent à un ensemble limité de relations spatiales symboliques (à gauche de, au dessus de, etc.).

Les approches quantitatives regroupent les méthodes qui visent à capturer précisément le positionnement spatial de régions les unes par rapport aux autres. Dans ce contexte, les méthodes basées sur la logique floue et la morphologie mathématique sont fréquemment utilisées (Bloch, 1999). Ces méthodes produisent un « paysage flou » pour chaque direction considérée, et permettent d'obtenir une évaluation riche et intuitivement pertinente d'une relation spatiale spécifique entre deux objets. On trouve des applications du modèle des paysages flous dans différents domaines tels le raisonnement spatial dans les images médicales (Bloch et Ralescu, 2003; Colliot et al., 2006), la reconnaissance d'écriture manuscrite (Delaye et Anquetil, 2011). Cependant, la combinaison de tels paysages pour produire une représentation globale de la structure spatiale entre régions n'est pas évidente. Une autre approche quantitative est basée sur le modèle des histogrammes d'angles (Miyajima et Ralescu, 1994) et de forces (Matsakis et Wendling, 1999). Ce modèle rend possible la manipulation des objets composés de plusieurs régions déconnectées, et permet de résumer leur position relative le long de toutes les directions au sein d'une même représentation.

En se basant sur le modèle des histogrammes de forces, nous proposons dans cet article, qui est une version étendue des travaux présentés dans (Clément et al., 2015b) (ainsi qu'une version francophone de travaux présentés dans (Clément et al., 2015a)), un descripteur homogène de formes et de relations spatiales adapté à la reconnaissances d'objets représentés dans des images en couleurs. Il s'agit notamment d'une extension des travaux de (Garnier

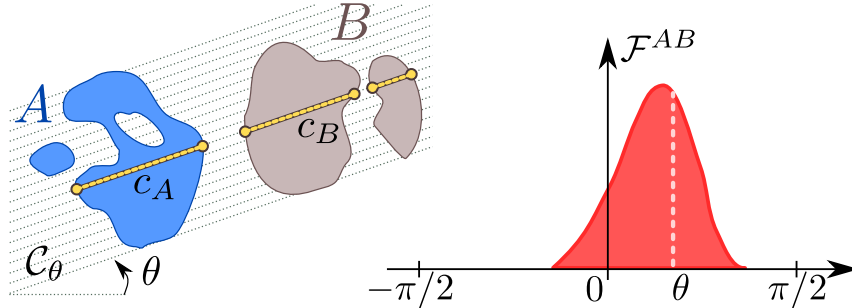


FIG. 1 – Illustration du calcul d'un histogramme de forces. La force d'attraction $\mathcal{F}_{AB}(\theta)$ entre deux objets A et B dans une direction θ , est calculée en considérant toutes les coupes longitudinales C_A et C_B qui balayent les deux objets dans cette direction.

et al., 2012) où ce descripteur avait été proposé pour des images en niveaux de gris. Dans notre extension, les objets sont décomposés en couches structurales par le biais d'une stratégie couplant l'algorithme de segmentation d'images Mean Shift à l'algorithme de classification non-supervisée K-Means. Cette première contribution permet d'extraire de manière pertinente les régions d'intérêt des objets dans des images en couleurs. Notre seconde contribution est de proposer différentes stratégies pour la comparaison de ces descripteurs en se basant sur une mise en correspondance des régions composant les objets lors du calcul d'une mesure de distance. Ces deux contributions permettent d'améliorer la reconnaissance d'objets structurés à partir d'images complexes.

La suite de cet article est organisée comme suit. En section 2, nous présentons la méthodologie proposée pour la reconnaissance d'objets dans des images en couleurs se basant sur les relations spatiales. Les validations expérimentales de cette méthodologie sur plusieurs bases d'images sont présentées dans la section 3. Enfin, des conclusions et perspectives sont proposées en section 4.

2 Méthodologie

Dans un premier temps, nous présentons brièvement le modèle des histogrammes de forces. Ensuite, nous décrivons notre méthodologie d'extraction des régions d'intérêt d'un objet dans une image en couleurs. Nous présentons ensuite le descripteur de formes et de relations spatiales basé sur le calcul d'un histogramme de forces entre chaque couple de régions issues de l'étape d'extraction. Enfin, nous proposons différentes stratégies de comparaison et de mise en correspondance de tels descripteurs pour leur utilisation dans des processus de reconnaissance et de classification d'objets.

2.1 Histogrammes de forces

Un histogramme de forces permet d'évaluer les relations spatiales directionnelles existant entre deux objets binaires représentés dans une image. Il s'agit d'un histogramme circulaire le

Description d'objets à partir des relations spatiales entre régions

long des directions d'angles $\theta \in [0, 2\pi[$. Ainsi, pour deux objets binaires A et B , chaque valeur de l'histogramme correspond à un angle θ , et mesure la valeur de vérité de la proposition « A est dans la direction θ de B », qui peut être vue comme une relation binaire floue $\mathcal{R}_\theta(A, B)$. L'ensemble de l'histogramme résume de manière globale le positionnement spatial de A par rapport à B , dans toutes les directions. Le modèle des histogrammes de forces est naturellement invariant par translation car les objets sont manipulés indépendamment de leur localisation dans l'image. Une illustration du calcul d'un histogrammes de forces entre deux objets A et B est présentée dans la Figure 1. Pour plus d'informations sur le fonctionnement et les propriétés du modèle des histogrammes de forces, voir (Matsakis et Wendling, 1999).

2.2 Extraction des régions d'intérêt des objets

Afin de décrire la structure spatiale interne d'un objet complexe présent dans une image, il est tout d'abord nécessaire de le décomposer en différentes sous-parties ou régions d'intérêt. Nous effectuons cette décomposition à l'aide d'un algorithme de segmentation, qui consiste à partitionner une image en différentes régions connexes selon différents critères. Cependant, dans notre contexte, les différentes sous-parties d'intérêt d'un objet peuvent être elles-mêmes composées de plusieurs régions déconnectées. Par conséquent, nous associons notre stratégie de segmentation à un algorithme de classification non-supervisée permettant de reconstruire de manière pertinente les différentes sous-parties de l'objet étudié.

2.2.1 Segmentation de l'objet en sous-parties

La segmentation d'images constitue un vaste champ de recherche à part entière, et l'efficacité des méthodes existantes reste bien souvent fortement dépendante de l'application envisagée. Dans ces travaux, nous avons choisi d'appliquer l'algorithme de segmentation Mean Shift (Comaniciu et Meer, 2002) qui est adapté à la manipulation d'objets complexes dans des images en couleurs, où les frontières entre les régions ne sont pas toujours bien définies. Le Mean Shift est un algorithme issu de l'apprentissage automatique, qui considère l'espace des caractéristiques comme une densité de probabilité, et qui cherche à en estimer les modes. Nous résumons ici son fonctionnement pour la segmentation d'images en couleurs. Celles-ci sont tout d'abord converties dans l'espace $L^*u^*v^*$ (luminance et chrominance) qui permet de répartir les couleurs de manière plus uniforme par rapport à la perception humaine. Pour chaque pixel, l'algorithme construit un voisinage dont il calcule la valeur moyenne. L'algorithme se déplace alors vers cette valeur moyenne et ce processus est itéré jusqu'à convergence, c'est-à-dire jusqu'à ce que tous les pixels soient assignés aux valeurs moyennes obtenues. Le Mean Shift produit une partition de l'image en R régions présentant des propriétés colorimétriques homogènes, et supposées délimiter les différentes régions d'intérêt composant l'image. Un résultat de segmentation d'un objet dans une image par l'algorithme Mean Shift est présenté en Figure 2 (b).

2.2.2 Reconstruction des sous-parties de l'objet

Le résultat de segmentation obtenu après application de l'algorithme Mean Shift produit un nombre de régions qui n'est pas connu par avance. De plus, toutes les régions sont composées uniquement de pixels connexes. Par exemple, en Figure 2 (b), la zone orange située sur les ailes

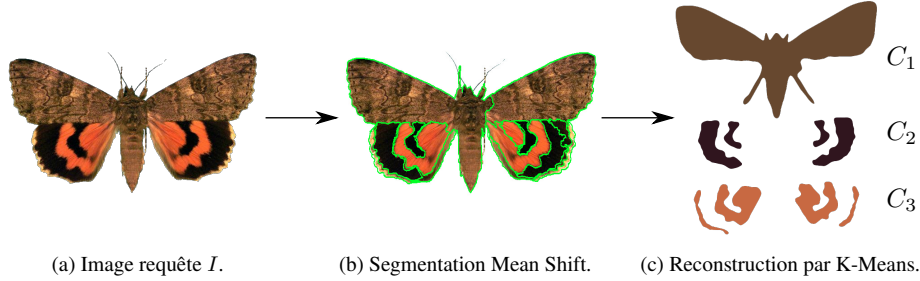


FIG. 2 – Illustration de la stratégie d'extraction des régions d'intérêt des objets.

du papillon est répartie en deux régions déconnectées, alors que d'un point de vue sémantique, nous souhaiterions qu'elles forment une seule et même sous-partie de l'objet. En ce sens, nous appliquons alors l'algorithme de classification non-supervisée K-Means (MacQueen, 1967) sur les pixels de l'image auxquels nous avons affecté la valeur d'intensité moyenne obtenue à l'issue de la segmentation Mean Shift. L'algorithme K-Means est initialisé avec N graines afin de produire un regroupement des régions de l'image en N couches structurelles, correspondant bien à la reconstruction des régions segmentées déconnectées. Dans la Figure 2 (c), les régions formées par l'étape de segmentation ont été reconstruites afin d'obtenir un total de 3 couches structurelles de l'objet.

2.3 Décomposition en histogrammes de forces

Une fois que l'objet représenté dans une image en couleurs a été décomposé en un ensemble de N régions d'intérêt C_1, \dots, C_N , nous proposons de calculer un descripteur spatial conformément à ce qui a été exposé dans (Garnier et al., 2012). Le principe de ce descripteur consiste à calculer un histogramme de forces $\mathcal{F}^{C_i C_j}$ entre chaque couple de régions (C_i, C_j) issues de l'étape d'extraction des régions d'intérêt. Lorsque $i = j$, l'histogramme de forces calculé correspond à une région avec elle-même : il s'agit d'une description de la forme de la région. Les histogrammes de forces calculés pour $i < j$ correspondent à une information sur le positionnement spatial relatif des régions, les unes par rapport aux autres. L'ensemble de ces histogrammes de forces constitue alors ce que nous appelons un descripteur *FHD* (pour *Force Histogram Decomposition*) :

$$\underbrace{\{\mathcal{F}^{C_i C_i}\}_{\forall i \in \{1..N\}}}_{\text{formes}} \cup \underbrace{\{\mathcal{F}^{C_i C_j}\}_{\forall (i,j) \in \{1..N\}^2, i < j}}_{\text{relations spatiales}} \quad (1)$$

Un descripteur FHD est donc composé de $N(N + 1)/2$ histogrammes de forces, et peut être représenté de manière matricielle, comme l'illustre la Figure 3. Cette matrice triangulaire est composée de N descripteurs de formes (sur la diagonale) et de $N(N - 1)/2$ descripteurs de relations spatiales (sur le triangle supérieur). Dans la suite, on dira d'un descripteur FHD composé de N régions qu'il est *de taille N*.

Compte tenu des propriétés d'invariance des histogrammes de forces, les descripteurs FHD sont naturellement invariants face à la *translation*, ainsi qu'à la *mise à l'échelle* si les his-

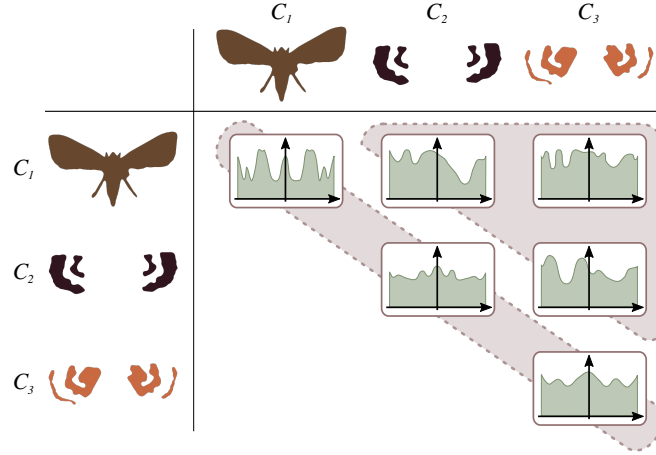


FIG. 3 – Illustration du descripteur FHD : un histogramme de forces est calculé entre chaque couple de sous-parties (C_i, C_j) issues de l'étape de décomposition, regroupant des descripteurs de formes et de relations spatiales.

togrammes de forces sont normalisés. Une estimation de la *rotation* peut être obtenue en effectuant des décalages circulaires des histogrammes de forces. De plus, ils proposent des propriétés symétriques intéressantes, ce qui permet notamment de ne pas avoir à calculer les histogrammes de forces pour le cas $i > j$ (triangle inférieur).

2.4 Comparaison des descripteurs FHD

Dans le but d'utiliser les descripteurs FHD dans un processus de reconnaissance d'objets ou de classification, il est important de définir une stratégie de comparaison de ces descripteurs. Il s'agit donc de construire une mesure de distance adaptée pour de tels descripteurs.

2.4.1 Mesure de distance

Étant donné deux descripteurs FHD de même taille N (correspondant au nombre de régions d'intérêt pour chaque objet), une manière de les comparer consiste à calculer la distance entre chacun de leurs histogrammes de forces pris deux-à-deux. Il existe de nombreuses mesures de distance entre histogrammes (Cha et Srihari, 2002). En se basant sur les résultats de (Garnier et al., 2012), nous avons choisi d'utiliser la distance du χ^2 pour comparer les histogrammes de forces deux-à-deux. La distance entre deux histogrammes de forces \mathcal{F}_A et \mathcal{F}_B le long de θ_{max} directions est donnée par :

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

Comme indiqué précédemment, le descripteur matriciel FHD comprend à la fois des descripteurs de formes (diagonale) et des descripteurs de relations spatiales (triangle supérieur). Il

paraît donc méthodologiquement pertinent de considérer séparément la distance entre ces deux types d'information. Ainsi, la distance globale entre deux descripteurs FHD Q et T peut être calculée de la manière suivante :

$$\mathcal{D}_\alpha(Q, T) = \alpha \mathcal{D}_{shape}(Q, T) + (1 - \alpha) \mathcal{D}_{spatial}(Q, T), \quad (3)$$

où α est un nombre réel entre 0 et 1 permettant d'affecter un poids variable à la distance \mathcal{D}_{shape} entre les descripteurs de formes et à la distance $\mathcal{D}_{spatial}$ entre les descripteurs de relations spatiales, ces distances étant définies de cette façon :

$$\mathcal{D}_{shape}(Q, T) = \frac{1}{N} \sum_{i=1}^N d_{\chi^2}(\mathcal{F}^{C_i C_i}(Q), \mathcal{F}^{C_i C_i}(T)), \quad (4)$$

$$\mathcal{D}_{spatial}(Q, T) = \frac{2}{N(N-1)} \sum_{i=1}^N \sum_{j=i+1}^N d_{\chi^2}(\mathcal{F}^{C_i C_j}(Q), \mathcal{F}^{C_i C_j}(T)). \quad (5)$$

Une telle pondération des distances $\mathcal{D}_{spatial}$ et \mathcal{D}_{shape} permet notamment d'éviter que les descripteurs de relations spatiales ne deviennent prédominants par rapport aux descripteurs de formes lorsque $N > 3$ (i.e., il y a alors plus d'histogrammes de forces pour les relations spatiales que pour les formes).

2.4.2 Mise en correspondance

La mesure de distance présentée ci-avant n'a de sens que si les régions d'intérêt structurant les deux objets comparés sont correctement alignées dans les deux matrices. En effet, il est possible que les régions issues de l'étape de décomposition pour deux objets ne soient pas ordonnées de la même façon, même si ceux-ci sont très similaires. Ce cas est notamment susceptible d'apparaître du fait que nous manipulons des images en couleurs : il n'y a alors pas d'ordre naturel entre les sous-parties des objets, celles-ci étant associées à des valeurs d'intensité dans l'espace RVB (rouge, vert, bleu). De ce fait, la distance entre les deux descripteurs risque d'être anormalement élevée, car calculée entre des histogrammes de forces qui ne correspondent pas deux à deux. Par conséquent, pour comparer des descripteurs FHD, il est nécessaire d'avoir recours à une stratégie de mise en correspondance des régions composant les objets, en amont du calcul de distance. Ce problème est de nature combinatoire : pour deux objets décomposés en N régions, il existe $N!$ possibilités d'appariement de leurs régions. Nous proposons ici d'employer des heuristiques se basant sur des hypothèses concernant les régions des objets pour réduire la complexité du problème.

Alignement par luminance La première stratégie consiste à partir de l'hypothèse que les régions correspondantes entre deux objets seront similaires en termes de colorimétrie. Afin de disposer d'un ordre total entre les couleurs associées aux sous-parties des objets, nous calculons leurs valeurs de luminance respectives dans l'espace $L^*u^*v^*$ (luminance et chrominance). Pour chaque objet, les régions sont alors triées par valeur de luminance. La Figure 4 (a) illustre le principe de cette stratégie.

Description d'objets à partir des relations spatiales entre régions

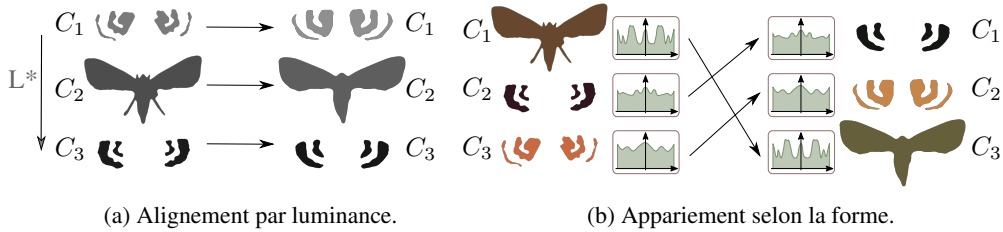


FIG. 4 – Illustration des deux stratégies de mise en correspondance proposées.

Appariement selon la forme La seconde stratégie part de l’hypothèse que les régions correspondantes entre deux objets seront similaires du point de vue de leurs formes. La Figure 4 (b) illustre ce principe : chaque sous-partie est appariée en fonction des histogrammes de forces décrivant la forme (c’est-à-dire ceux situés sur la diagonale du descripteur). Pour mettre en place cette stratégie, il est possible de l’assimiler à un problème d’optimisation dont l’objectif est de minimiser la distance globale $\mathcal{D}_{shape}(Q, T)$ entre les descripteurs de formes respectifs des deux objets. Nous proposons alors deux approches. L’une consiste à énumérer les $N!$ possibilités d’appariement entre les régions des deux objets : il s’agit d’une solution « optimale », qui est cependant inappropriée pour des descripteurs FHD composés d’un nombre élevé de régions. Une approximation de cette solution consiste à mettre en correspondance les régions successivement en faisant à chaque fois le meilleur choix possible : il s’agit d’une stratégie « gloutonne ». D’un point de vue algorithmique, cette approximation peut être considérée comme un bon compromis nécessitant une complexité quadratique en fonction de N .

Le choix d’une stratégie de mise en correspondance est fortement dépendant de l’application envisagée et chacune présente des avantages et des inconvénients. D’un côté, la stratégie d’alignement par luminance est relativement simple et peut être considérée comme une étape de pré-traitement, cependant, elle est vouée à échouer si la luminance n’est pas une caractéristique discriminante (par exemple, si deux objets ont la même structure mais des couleurs différentes). D’un autre côté, la stratégie d’appariement selon la forme paraît plus appropriée, mais implique d’être appliquée à chaque comparaison de deux descripteurs. De plus, cette stratégie est également plus disposée à commettre des erreurs, notamment lorsque certaines régions peuvent être assimilées à du bruit.

3 Validations expérimentales

3.1 Jeux de données

Dans le cadre de nos expérimentations, deux bases d’images représentant des objets en couleurs ont été utilisées. La première est un ensemble d’images de papillons, et la deuxième regroupe des images de motos. Nous avons choisi de valider l’approche proposée sur des jeux de données spécifiques où les objets d’intérêt sont représentés sur un arrière-plan homogène, évitant ainsi les difficultés liées à la détection d’objets.



FIG. 5 – Quelques images de papillons issues du jeu de données PEALE.



FIG. 6 – Quelques images de motos issues du jeu de données MOTOS.

Base d'images PEALE Les papillons sont un cas typique d'objets où la forme ainsi que la disposition spatiale des motifs sur les ailes constituent des caractéristiques particulièrement discriminantes pour distinguer les espèces. La base d'images PEALE¹ est composée de 318 images couleur, présentant chacune un papillon vu du dessus. Les papillons appartiennent à des espèces différentes, divisant le jeu de données en un total de 28 classes non équiréparties, et constituant notre vérité terrain pour les tâches de reconnaissance et de classification. Les images de ce jeu de données ont une taille moyenne de 640 pixels de large. Quelques images issues du jeu de données PEALE sont présentées dans la Figure 5.

Base d'images MOTOS De la même façon que pour les papillons, les motos sont un cas intéressant où la forme et la disposition spatiale des composants tels que les roues, le carénage ou le moteur permettent naturellement de distinguer les différents types de motos. La base d'images MOTOS est un sous-ensemble d'une plus grande base issue de l'Université Caltech², composée d'images couleur de motos variées, toutes vues de profil. Le sous-ensemble d'images utilisé dans nos travaux est composé de 265 images présentant un fond homogène, et d'une taille typique de 500 pixels de large. Afin de disposer d'une vérité terrain, nous avons manuellement annoté les images du jeu de données. Nous les avons réparties en 5 catégories : *cross*, *customs*, *sportives*, *roadsters* et *cyclomoteurs*. Quelques images issues de ce jeu de données sont présentées dans la Figure 6.

3.2 Protocole expérimental

Pour chaque jeu de données, nous disposons d'une vérité terrain, c'est-à-dire qu'une classe thématique (ou label) est assignée à chaque image. Afin d'éviter un quelconque biais dans nos expérimentations, nous utilisons un protocole de validation croisée de type « *leave-one-out* » : chaque image d'objet (échantillon de test) est comparée au reste du jeu de données

1. Academy of Natural Sciences, Philadelphia. <http://clade.ansp.org/entomology>

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

Description d'objets à partir des relations spatiales entre régions

Décomposition / N	2	3	4	5	6
Niveaux de gris	26,1	30,7	31,9	37,4	40,8
Mean Shift $(h_s, h_r, M) = (8, 4, 100)$	44,7	42,8	39,3	41,2	43,7
Mean Shift $(h_s, h_r, M) = (8, 8, 100)$	46,9	39,0	36,2	38,4	35,2
Mean Shift $(h_s, h_r, M) = (8, 12, 100)$	41,2	35,8	34,9	29,6	34,9

TAB. 1 – Taux de reconnaissance global T_R pour le jeu de données PEALE pour différentes méthodes de décomposition des objets, en faisant varier le nombre N de sous-parties. Les triplets (h_s, h_r, M) correspondent aux paramètres utilisés pour la segmentation Mean Shift.

Décomposition / N	2	3	4	5	6
Niveaux de gris	65,2	59,1	58,1	55,6	63,1
Mean Shift $(h_s, h_r, M) = (8, 4, 100)$	69,8	71,3	63,4	63,8	62,6
Mean Shift $(h_s, h_r, M) = (8, 8, 100)$	67,9	64,9	62,3	60,8	58,9
Mean Shift $(h_s, h_r, M) = (8, 12, 100)$	66,8	63,8	67,2	56,6	55,8

TAB. 2 – Taux de reconnaissance global T_R pour le jeu de données MOTOS pour différentes méthodes de décomposition des objets, en faisant varier le nombre N de sous-parties. Les triplets (h_s, h_r, M) correspondent aux paramètres utilisés pour la segmentation Mean Shift.

(échantillon d'apprentissage), ceci successivement pour toutes les images. La classification de l'échantillon de test se fait par recherche du plus proche voisin : on assigne à l'image requête la classe de l'image la plus proche dans l'espace des caractéristiques engendré par l'ensemble des descripteurs FHD de l'échantillon d'apprentissage. Nous calculons alors le taux de reconnaissance global T_R .

Les histogrammes de forces sont calculés avec une force d'attraction constante et sur un total de 180 directions discrétisées, balayant l'intervalle $[0, 2\pi[$ par pas réguliers de 2 degrés. Dans l'équation 3, nous avons fixé $\alpha = 0.5$ par défaut, afin de donner un poids équivalent aux descripteurs de formes et aux descripteurs de relations spatiales lors des calculs de distances.

3.3 Résultats expérimentaux

Taux de reconnaissance Le Tableau 1 et le Tableau 2 présentent les taux de reconnaissance obtenus pour la classification des jeux de données PEALE et MOTOS selon le protocole présenté précédemment. Les résultats sont présentés pour différentes méthodes de décomposition des objets, en faisant varier le nombre N de sous-parties des objets. La première ligne présente les résultats pour la version originale des FHD introduite par (Garnier et al., 2012) où les objets sont décomposés selon leurs niveaux de gris. Les lignes suivantes montrent les résultats obtenus pour notre extension des descripteurs FHD où les objets sont issus d'images en couleurs. Différents paramètres sont utilisés pour l'algorithme de segmentation Mean Shift. Le paramètre h_s correspond à la fenêtre spatiale utilisée par l'algorithme, et a été fixé empiriquement à $h_s = 8$ en fonction de la taille des images. Le paramètre h_r correspond à l'intervalle d'in-

Appariement / N	2	3	4	5	6
Luminance	46,9	42,8	39,3	41,2	43,7
Formes (glouton)	43,7	42,1	33,0	34,0	31,4
Formes (optimal)	44,0	40,9	33,3	31,8	34,3

TAB. 3 – Taux de reconnaissance global T_R pour le jeu de données PEALE pour les différentes stratégies d'appariement des descripteurs.

Appariement / N	2	3	4	5	6
Luminance	69,8	71,3	67,2	63,8	62,6
Formes (glouton)	68,7	64,9	61,9	61,5	58,5
Formes (optimal)	68,7	63,8	63,0	62,3	60,8

TAB. 4 – Taux de reconnaissance global T_R pour le jeu de données MOTOS pour les différentes stratégies d'appariement des descripteurs.

tensité parmi lequel les pixels d'un voisinage sont considérés comme appartenant à une même région. Ainsi, une valeur faible de ce paramètre, ici $h_r = 4$, produit une sur-segmentation de l'image, conservant notamment les détails de texture. À l'inverse, une valeur plus élevée, $h_r = 12$ produit des régions plus grandes et plus hétérogènes. Enfin, le paramètre M correspond à la taille minimum des régions, fixée à $M = 100$ selon le même raisonnement.

On constate que les résultats obtenus pour l'extension des descripteurs FHD aux images en couleurs sont dans l'ensemble supérieurs à ceux obtenus pour la décomposition en niveaux de gris proposée initialement. Cela confirme notre hypothèse selon laquelle une méthode de décomposition prenant en compte simultanément l'information spatiale et colorimétrique dans les images (par le biais d'une approche de segmentation) permet de construire des régions d'intérêt plus représentative de la structure des objets de ces jeux de données. Par ailleurs, nous pouvons également constater que les résultats semblent plus stables en adoptant une stratégie de sur-segmentation (*i.e.*, valeur de h_r faible), en particulier lorsque les objets sont décomposés en un nombre de sous-parties de plus en plus élevé.

Mise en correspondance Le Tableau 3 et le Tableau 4 présentent les taux de reconnaissance obtenus sur les jeux de données PEALE et MOTOS, pour les différentes stratégies de mise en correspondance proposées en Section 2.4.2. La première ligne du tableau montre les résultats obtenus pour la stratégie d'alignement des régions par luminance. Les deux lignes suivantes présentent les résultats obtenus pour la stratégie de mise en correspondance en fonction des formes des régions, la première par appariement glouton, et la seconde par appariement optimal.

Pour la base d'images PEALE, on constate que la stratégie d'alignement des régions par luminance fournit de meilleurs résultats pour toutes les valeurs de N . Cela peut être dû au fait que pour la plupart des espèces de papillons du jeu de données, la couleur des régions d'intérêt constitue la caractéristique la plus discriminante. Pour la base d'images MOTOS, la

Description d'objets à partir des relations spatiales entre régions

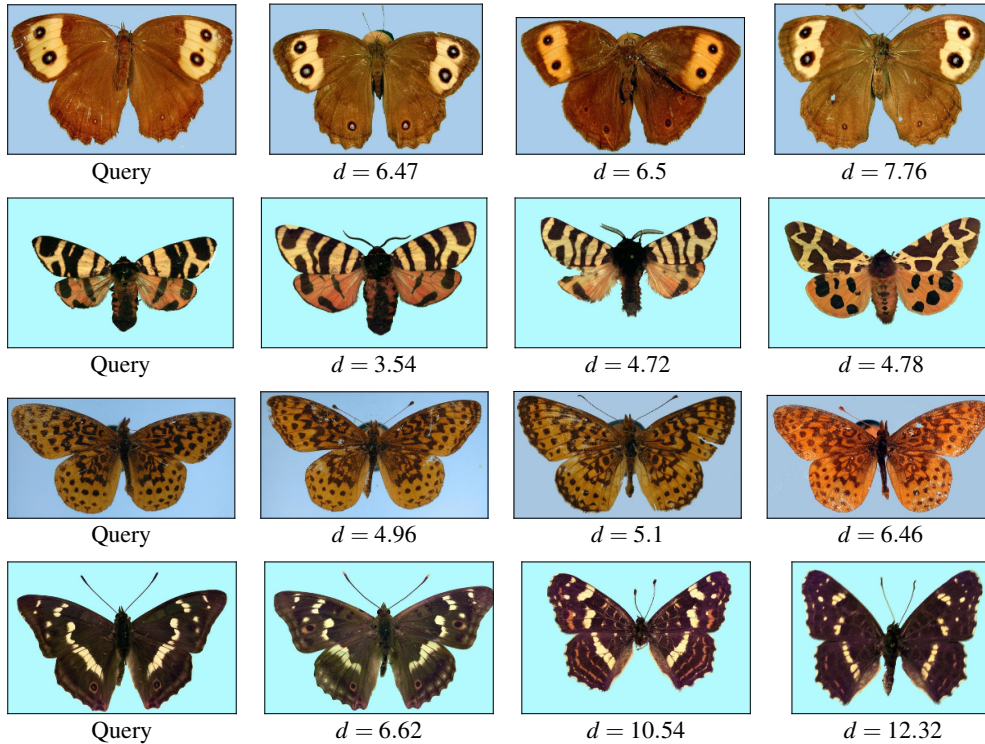


FIG. 7 – Quelques résultats représentatifs de recherche d'objets similaires pour la base d'image PEALE. Pour chaque ligne, l'image de gauche représente l'image requête, par ordre décroissant de similarité.

stratégie d'appariement par la forme permet d'obtenir des résultats comparables à l'alignement par luminance. En effet, dans ce jeu de données, les motos se distinguent principalement par la structure et la forme de leurs sous-parties, plutôt que par leurs couleurs (par exemple, de nombreuses motos correspondent à des modèles identiques mais présentent des couleurs différentes).

Globalement, pour les deux jeux de données, nous pouvons également constater que pour la stratégie d'appariement sur la forme, l'approche gloutonne produit des résultats comparables, et parfois même supérieurs, à ceux de l'approche optimale, mais avec un coût de calcul beaucoup plus faible. Le fait que la stratégie gloutonne puisse fournir de meilleurs résultats que la stratégie optimale peut être dû au fait que cette dernière n'est optimale qu'au sens de la distance, elle n'implique pas forcément une correspondance parfaite entre les sous-parties (qui n'existe pas forcément, en particulier lors de la comparaison d'objets très différents). En plus de ces résultats globaux, nous avons étudié au niveau de la reconnaissance par classe le comportement des différentes stratégies de mise en correspondance. Pour plus de détails sur ces résultats, les lecteurs pourront se référer à (Clément et al., 2015a).



FIG. 8 – Quelques résultats représentatifs de recherche d'objets similaires pour la base d'image MOTOS. Pour chaque ligne, l'image de gauche représente l'image requête, par ordre décroissant de similarité.

Évaluation qualitative La Figure 7 et la Figure 8 présentent quelques résultats représentatifs de recherche d'objets similaires sur les jeux de données PEALE et MOTOS, en utilisant la stratégie de mise en correspondance basée sur la luminance pour PEALE ($N = 3$ sous-parties), et la mise en correspondance gloutonne basée sur la forme pour MOTOS ($N = 4$ sous-parties). D'après ces résultats visuels, il apparaît que les descripteurs FHD couplés à ces différentes stratégies d'appariement permettent de reconnaître de manière pertinente ce type d'objets structurés.

Influence du paramètre α Pour compléter cette étude expérimentale, nous avons également étudié l'impact du paramètre α modélisant le poids donné respectivement aux descripteurs de formes par rapport aux descripteurs de relations spatiales (voir Équation 3). Au lieu de fixer ce paramètre de manière « globale » ($\alpha = 0.5$) comme cela a été réalisé dans les expérimentations précédentes, nous avons évalué le meilleur poids « local » à attribuer à chaque classe des jeux de données. Nous avons exhaustivement calculé les taux de reconnaissance « classe-par-classe » en faisant varier $\alpha \in [0, 1]$ successivement avec un pas de 0.02. Pour chaque classe i , nous enregistrons les valeurs α_i conduisant aux meilleurs taux de reconnaissance pour cette classe

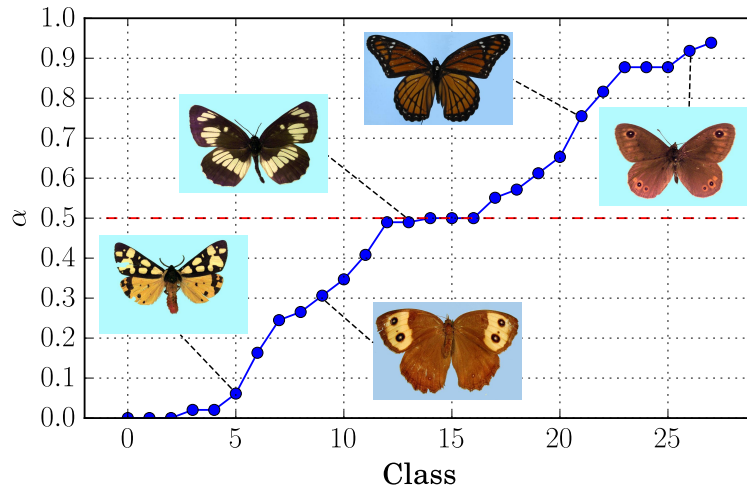


FIG. 9 – Synthèse des meilleurs poids α obtenus pour chaque classe de la base d'images PEALE (avec $N = 3$). Les classes sont triées selon les valeurs croissantes de α . Des valeurs faibles indiquent un poids plus important accordé aux relations spatiales, tandis que des valeurs élevées indiquent un poids plus important accordé à la forme des régions.

en particulier. La Figure 9 présente les résultats de cette étude sur la base d'images PEALE, avec quelques exemples d'objets appartenant à des classes représentatives. À partir de cette figure, nous pouvons observer que suivant la classe considérée, les descripteurs de relations spatiales ou de formes peuvent être des caractéristiques plus ou moins discriminantes. Parmi les 28 classes de ce jeu de données, environ un tiers semble plus orienté vers l'information spatiale (valeurs de α faibles), un autre tiers plutôt vers des informations de forme (valeurs de α élevées), et le reste est relativement équilibré entre les deux.

4 Conclusion

Dans cet article, nous avons introduit une nouvelle approche pour la reconnaissance d'objets à partir d'images en couleurs. La principale originalité de cette approche repose sur une extension d'un descripteur d'objets qui intègre dans une représentation homogène des informations de formes et de relations spatiales entre les différentes régions d'intérêt composant les objets. Les premiers résultats obtenus sur deux bases d'images complexes montrent que de telles caractéristiques sont adaptées à la reconnaissance d'objets structurés au sein d'images en couleurs.

Ces travaux ont ouvert la voie à différentes perspectives. Une perspective à court terme consistera à envisager une décomposition des objets en un nombre variable de sous-parties. Ceci pourrait notamment être effectué par le biais d'une approche de segmentation hiérarchique des images. Ce type de segmentation adaptative permet de former des régions d'in-

térêt à différents niveaux d'échelle, résultant en une représentation des images sous la forme d'arbres de partition. Par ailleurs, une telle décomposition impliquerait alors d'être capable de comparer des descripteurs composés d'un nombre variable d'histogrammes de forces. Enfin, des travaux plus théoriques sur la modélisation et l'enrichissement du pouvoir de description des histogrammes de forces sont également envisagés.

Références

- Andreopoulos, A. et J. K. Tsotsos (2013). 50 Years of object recognition : Directions forward. *Computer Vision and Image Understanding* 117(8), 827–891.
- Belarte, B., C. Wemmer, G. Forestier, M. Grizonnet, et C. Weber (2013). Learning fuzzy rules to characterize objects of interest from remote sensing images. In *Proceedings of the IEEE Geoscience and Remote Sensing Symposium – IGARSS 2013*, pp. 2986–2989.
- Bloch, I. (1999). Fuzzy relative position between objects in image processing : A morphological approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 21(7), 657–664.
- Bloch, I. et A. Ralescu (2003). Directional relative position between objects in image processing : A comparison between fuzzy approaches. *Pattern Recognition* 36(7), 1563–1582.
- Cha, S.-H. et S. N. Srihari (2002). On measuring the distance between histograms. *Pattern Recognition* 35(6), 1355–1370.
- Clément, M., M. Garnier, C. Kurtz, et L. Wendling (2015a). Color object recognition based on spatial relations between image layers. In *Proceedings of the International Conference on Computer Vision Theory and Applications – VISAPP 2015*, pp. 427–434.
- Clément, M., C. Kurtz, et L. Wendling (2015b). Descripteurs de relations spatiales entre régions structurelles pour la reconnaissance d'objets en couleurs. In *Proceedings of the Atelier Fouille de Données Complexes, Journées Francophones Extraction et Gestion des Connaissances (EGC 2015) – FDC 2015*, pp. 25–36.
- Cohn, A. G., B. Bennett, J. Gooday, et N. M. G. Gotts (1997). Qualitative spatial representation and reasoning with the region connection calculus. *GeoInformatica* 1(3), 275–316.
- Colliot, O., O. Camara, et I. Bloch (2006). Integration of fuzzy spatial relations in deformable models - Application to brain MRI segmentation. *Pattern Recognition* 39(8), 1401–1414.
- Comaniciu, D. et P. Meer (2002). Mean shift : A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24(5), 603–619.
- Delaye, A. et E. Anquetil (2011). Fuzzy relative positioning templates for symbol recognition. In *Proceedings of the IEEE International Conference on Document Analysis and Recognition – ICDAR 2011*, pp. 1220–1224.
- Egenhofer, M. J. (1989). A formal definition of binary topological relationships. In *Foundations of Data Organization and Algorithms*, Volume 367 of *Lecture Notes in Computer Science*, pp. 457–472.
- Freeman, J. (1975). The modelling of spatial relations. *Computer Graphics and Image Processing* 4(2), 156–171.

Description d'objets à partir des relations spatiales entre régions

- Garnier, M., T. Hurtut, et L. Wendling (2012). 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*, pp. 1–7.
- Ilea, D. E. et P. F. Whelan (2011). Image segmentation based on the integration of colour–texture descriptors – a review. *Pattern Recognition* 44(10), 2479–2501.
- Inglada, J. et J. Michel (2009). Qualitative spatial reasoning for high-resolution remote sensing image analysis. *IEEE Transactions on Geoscience and Remote Sensing* 47(2), 599–612.
- Jolion, J. M. (2001). *Feature Similarity*, pp. 122–162. Springer Verlag.
- MacQueen, J. B. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability – BSMSP 1967*, pp. 281–297.
- Matsakis, P. et L. Wendling (1999). A new way to represent the relative position between areal objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 21(7), 634–643.
- Miyajima, K. et A. Ralescu (1994). Spatial organization in 2D segmented images : Representation and recognition of primitive spatial relations. *Fuzzy Sets and Systems* 65(2), 225–236.

Summary

Classical object recognition methods from images usually rely on a statistical or structural description of the object content, summarizing different visual features such as contours, color or texture. Such features can then be used for the classification or the recognition of similar images. In this paper, we present a descriptor for complex objects represented in color images. The originality of this descriptor lies in an homogeneous representation of both shape attributes and spatial relations between structural regions extracted from objects using a robust decomposition method. We also propose different comparison strategies for this descriptor, based on specific matchings between structural regions of the objects. Results obtained on two datasets of color images suggest that the spatial relations between regions composing complex objects constitute interesting features for their description.