

Classification d'images à l'aide de noyaux sur graphes de segmentation

Zaïd HARCHAOU¹ & Francis BACH²

¹LTCI, Télécom Paris & CNRS
46 rue Barrault, 75634 Paris cedex 13 France

²CMM, École Nationale Supérieure des Mines
35 rue Saint-Honoré, 77305 Fontainebleau Cedex, France
zaid.harchaoui@enst.fr, francis.bach@mines.org

Résumé – Nous proposons une famille de noyaux entre images mesurant la similarité entre leurs graphes de segmentation. Les noyaux correspondent à l'appariement inexact de formes arborescentes présentes dans chacune des images. Notre approche résiste à l'instabilité du processus de segmentation tout en exploitant autant que possible la structure naturelle inhérente aux images. Le cadre des méthodes à base de noyaux nous permet de réaliser efficacement une large palette de tâches, telles que la classification semi-supervisée ou l'apprentissage de combinaison de noyaux.

Abstract – We propose a family of kernels between images, defined as kernels between their respective segmentation graphs. The kernels are based on soft matching of subtree-patterns of the respective graphs, leveraging the natural structure of images while remaining robust to the associated segmentation process uncertainty. Indeed, output from morphological segmentation is often represented by a labelled graph, each vertex corresponding to a segmented region, with edges joining neighboring regions. However, such image representations have mostly remained underused for learning tasks, partly because of the observed instability of the segmentation process and the inherent hardness of inexact graph matching with uncertain graphs. Our kernels count common virtual substructures amongst images, which enables to perform efficient supervised classification of natural images with a support vector machine. Moreover, the kernel machinery allows us to take advantage of recent advances in kernel-based learning: i) semi-supervised learning reduces the required number of labelled images, while ii) multiple kernel learning algorithms efficiently select the most relevant similarity measures between images within our family.

1 Introduction

Les images sont des objets vivant dans des espaces de grandes dimensions. La plupart des techniques actuelles contournent cette difficulté en commençant par réduire la dimension des données à travers une phase de pré-traitement réalisée *a priori*, qui permet d'extraire des caractéristiques pertinentes. Nous sommes ainsi partagés entre utiliser des méthodes évoluées sur des données très simplifiées après un pré-traitement complexe, ou des méthodes élémentaires sur des données restées proches des images originales.

Les réseaux de convolution [7] offrent une solution élégante en abordant ces deux facettes d'un seul coup. La manipulation de ces méthodes reste cependant délicate et difficile à analyser en termes statistiques. Les méthodes à base de noyaux comme les SVMs (*Support Vector Machines*) [13] reposent quant à elles sur l'optimisation d'un critère convexe. De manière surprenante, aller au-delà de l'approche commune consistant à appliquer un noyau gaussien à des caractéristiques globales demeure encore un défi d'actualité.

Nous proposons de mettre à profit la structure inhérente aux images en comparant deux images *via* un noyau d'appariement de sous-structures. Cette approche est inspirée des méthodes les plus récentes en bioinformatique, ou elle

s'est révélée très fructueuse [9] et représente maintenant l'état de l'art [12]. Une première application en vision artificielle a récemment été proposée dans [14]. Nous proposons de représenter les images par leurs graphes de segmentation. En effet, ils conservent l'essentiel de la structure des images, tout en offrant une représentation facilement maniable.

Cependant, le problème général de l'appariement inexact de graphes est NP-difficile, et celui de l'appariement de sous-graphes est NP-complet. Cette difficulté peut alors être surmontée en appariant plutôt des sous-structures virtuelles, construites par exemple à partir de promenades le long des graphes [11]. Notre approche consiste ici à définir des noyaux par appariement inexact de promenades arborescentes, en tirant parti du caractère planaire des graphes de segmentation. Le cadre des méthodes à base de noyaux permet alors d'aborder toute la palette des tâches d'apprentissage liées aux images, de la classification supervisée à l'indexation automatique.

2 Segmentation morphologique

La segmentation morphologique hiérarchique [10] est un outil puissant de traitement d'images permettant la segmentation rapide d'images pour un nombre de régions



FIG. 1 – Exemple d'image naturelle.

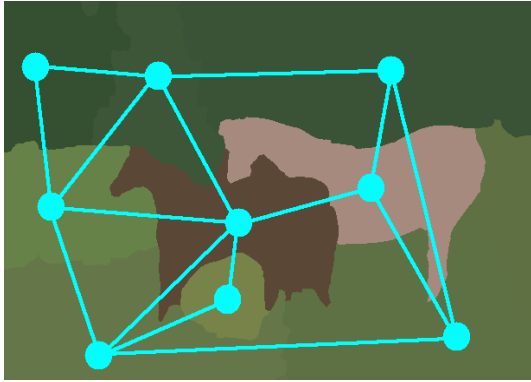


FIG. 2 – Mosaique de segmentation, avec la couleur médiane pour chaque région. Le graphe de segmentation est représenté en vert clair.

donné. Une partition de l'image est tout d'abord calculée à l'aide de la ligne de partage des eaux (LPE) à partir de l'image gradient. Puis les régions obtenues sont fusionnées par inondation pour atteindre le nombre de régions spécifié (100 régions dans nos expériences). Le résultat de la segmentation est alors un graphe planaire étiqueté, dont chaque sommet correspond à une région, et chaque arête relie deux régions adjacentes.

Les graphes de segmentation obtenus possèdent plusieurs caractéristiques cruciales, dont notre méthode peut tirer profit. Les images sont volontairement sur-segmentées, de manière à coller autant que possible à l'image initiale et à éviter les artefacts de segmentation. Pour chaque sommet du graphe, les sommets voisins peuvent être naturellement ordonnés du fait de la planarité du graphe. Enfin, typiquement les graphes sont creux (*sparse*); autrement dit les sommets possèdent en moyenne peu de sommets voisins.

3 Noyaux d'appariement de promenades

L'idée est de comparer deux images en comparant leurs graphes de segmentation respectifs. Dans la suite, pour deux images données les graphes de segmentation sont

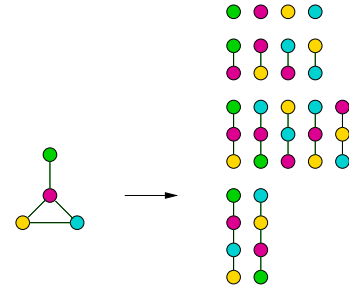


FIG. 3 – Chemins d'un graphe, regroupées par longueur. Les couleurs correspondent aux étiquettes des sommets.

notes \mathbf{G} et \mathbf{H} , les ensembles de sommets $\mathcal{S}_{\mathbf{G}}$ and $\mathcal{S}_{\mathbf{H}}$.

3.1 Noyau de base

La première brique des noyaux que nous proposons est le noyau $k(s, s')$ défini entre deux régions s et s' . Le noyau $k(\cdot, \cdot)$ est un noyau semi-défini positif s'appliquant aux deux descripteurs respectifs ℓ et ℓ' étiquetant les régions. Nous avons choisi d'utiliser les histogrammes locaux de couleurs $\ell \equiv P_{\ell}$ comme descripteurs des régions. Pour mesurer la similarité entre deux régions de graphes de segmentation, nous avons utilisé une variante du noyau dit du χ^2 entre histogrammes de couleur [6], qui tient compte des tailles des régions. Le noyau s'écrit $k(\ell, \ell') = \lambda A_{\ell}^{\gamma} A_{\ell'}^{\gamma} k_{\chi^2}(\ell, \ell')$, avec A_{ℓ} l'aire de la région d'étiquette ℓ , et $\gamma \in [0, 1]$ un paramètre réglant l'influence de la taille des régions.

3.2 Noyau d'appariement de promenades

Pour un graphe \mathbf{G} , une *promenade* est une suite de sommets voisins deux à deux, tandis qu'un *chemin* est une promenade dont tous les sommets sont distincts. La notion de promenade peut être comprise comme une relaxation de celle de chemin. La différence entre l'ensemble des chemins et des promenades d'un même graphe est illustrée dans Figure 3 et Figure 4. Alors que l'énumération des chemins d'un graphe conduit rapidement à des problèmes NP-difficiles [4], les promenades permettent quant à elles de développer des algorithmes tournant en temps polynomial.

Notons $\mathcal{P}_{\mathbf{G}}^p$ (resp. $\mathcal{P}_{\mathbf{H}}^p$) l'ensemble des promenades de longueur p de \mathbf{G} (resp. \mathbf{H}). Le noyau $k_{\mathcal{P}}^p(\mathbf{G}, \mathbf{H})$ entre promenades d'ordre p entre \mathbf{G} and \mathbf{H} est défini par

$$k_{\mathcal{P}}^p(\mathbf{G}, \mathbf{H}) = \sum_{\substack{(r_1, \dots, r_p) \in \mathcal{P}_{\mathbf{G}}^p \\ (s_1, \dots, s_p) \in \mathcal{P}_{\mathbf{H}}^p}} \prod_{i=1}^p k(\ell_{\mathbf{G}}(r_i), \ell_{\mathbf{H}}(s_i)).$$

La valeur du noyau peut être aisément calculée à l'aide d'un algorithme de programmation dynamique [5]. Soient $\mathcal{P}_{\mathbf{G}}^p(r)$ (resp. $\mathcal{P}_{\mathbf{H}}^p(s)$) l'ensemble des promenades de \mathbf{G} (resp. \mathbf{H}) demarrant en r (resp. s).

$$k_{\mathcal{P}}^p(\mathbf{G}, \mathbf{H}, r, s) = \sum_{\substack{(r_1, \dots, r_p) \in \mathcal{P}_{\mathbf{G}}^p(r) \\ (s_1, \dots, s_p) \in \mathcal{P}_{\mathbf{H}}^p(s)}} \prod_{i=1}^p k(\ell_{\mathbf{G}}(r_i), \ell_{\mathbf{H}}(s_i)).$$

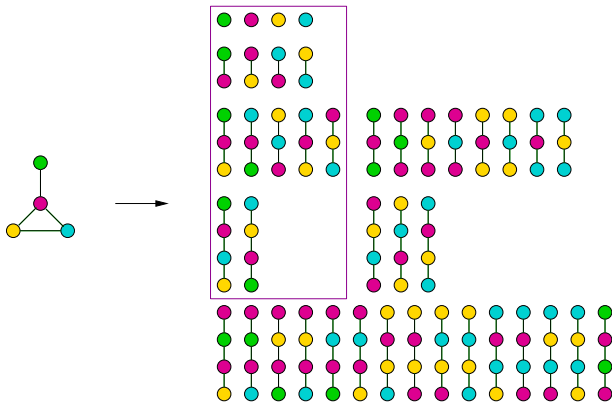


FIG. 4 – Promenades d’un graphe, regroupées par longueur. Les couleurs correspondent aux étiquettes des sommets. Les promenades situées à l’intérieur du cadre sont en réalité des chemins.

Soient $\mathcal{V}_{\mathbf{G}}(r)$ (resp. $\mathcal{V}_{\mathbf{H}}(s)$) l’ensemble des voisins de r dans \mathbf{G} (resp. de s dans \mathbf{H}). Les valeurs du noyau peuvent alors être calculées de manière récursive grâce la propriété ci-dessous.

$$k_{\mathcal{P}}^p(\mathbf{G}, \mathbf{H}, r, s) = k(\ell_{\mathbf{G}}(r), \ell_{\mathbf{H}}(s)) \sum_{\substack{r' \in \mathcal{V}_{\mathbf{G}}(r) \\ s' \in \mathcal{V}_{\mathbf{H}}(s)}} k_{\mathcal{P}}^{p-1}(\mathbf{G}, \mathbf{H}, r', s').$$

en initialisant à : $k_{\mathcal{P}}^1(\mathbf{G}, \mathbf{H}, r, s) = k(\ell_{\mathbf{G}}(r), \ell_{\mathbf{H}}(s))$.

Finalement, la valeur du noyau est donnée par :

$$k_{\mathcal{P}}^p(\mathbf{G}, \mathbf{H}) = \sum_{\substack{r \in \mathcal{V}_{\mathbf{G}} \\ s \in \mathcal{V}_{\mathbf{H}}}} k_{\mathcal{P}}^p(\mathbf{G}, \mathbf{H}, r, s).$$

3.3 Noyau d’appariement de promenades arborescentes

Pour capturer la topologie des images, nous proposons d’apparier des sous-structures virtuelles des graphes de segmentation : les promenades arborescentes. Une promenade arborescente α -aire $a_{\mathbf{G}}^{\alpha}$ d’un graphe \mathbf{G} est un arbre enraciné (*rooted subtree-pattern*), dont les sommets sont des sommets de \mathbf{G} , et tel que deux sommets voisins dans $a_{\mathbf{G}}^{\alpha}$ sont également voisins dans \mathbf{G} . Intuitivement, une **promenade** $a_{\mathbf{G}}^{\alpha}$ de \mathbf{G} est donc un arbre dont la structure de voisinage coïncide avec la structure de voisinage de \mathbf{G} , mais qui n’est pas forcément un vrai sous-arbre de \mathbf{G} .

Pour restreindre le nombre de promenades arborescentes considérées pour l’appariement, nous exploitons la planarité des graphes de segmentation. En effet, l’ensemble des voisins d’un sommet r peut être décomposé en intervalles $I \in \mathcal{I}_{\mathbf{G}}^{\alpha}(r)$ de largeur α de sommets voisins (cf. figure 6).

En sommant sur tous les intervalles I et J de même largeur, les noyaux d’appariement de promenades arborescentes $k_{\mathcal{A}}^{p,\alpha}(\mathbf{G}, \mathbf{H}, r, s)$ se calculent alors efficacement par programmation dynamique :

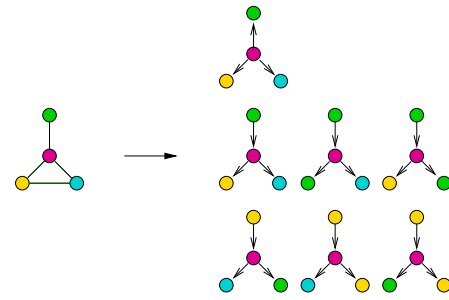


FIG. 5 – Exemples de promenades arborescentes

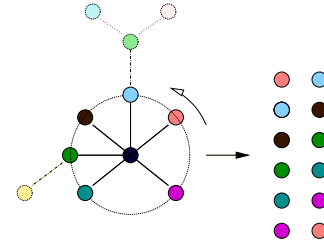


FIG. 6 – Intervalles de voisins de largeur 2

$$k_{\mathcal{A}}^{p,\alpha}(\mathbf{G}, \mathbf{H}, r, s) = k(\ell_{\mathbf{G}}(r), \ell_{\mathbf{H}}(s)) \times \sum_{\substack{I \in \mathcal{I}_{\mathbf{G}}^{\alpha}(r) \\ J \in \mathcal{I}_{\mathbf{H}}^{\alpha}(s)}} \prod_{\substack{r' \in I \\ s' \in J}} k_{\mathcal{A}}^{p-1,\alpha}(\mathbf{G}, \mathbf{H}, r', s').$$

où l’on somme sur I and J tels que $\text{card}(I)=\text{card}(J)$.

En définitive, le noyau d’appariement de promenades arborescentes α -aire de profondeur p entre deux graphes \mathbf{G} et \mathbf{H} vaut :

$$k_{\mathcal{A}}^{p,\alpha}(\mathbf{G}, \mathbf{H}) = \sum_{r \in \mathcal{S}_{\mathbf{G}}, s \in \mathcal{S}_{\mathbf{H}}} k_{\mathcal{A}}^{p,\alpha}(\mathbf{G}, \mathbf{H}, r, s).$$

Le noyau ainsi obtenu est **semi-défini positif**. De plus, pour deux graphes \mathbf{G} et \mathbf{H} , ayant respectivement $n_{\mathbf{G}}$ et $n_{\mathbf{H}}$ sommets, et de degrés maximaux respectifs $d_{\mathbf{G}}$ et $d_{\mathbf{H}}$, la complexité algorithmique du calcul du noyau pour des sous-arbres virtuels α -aire de profondeur maximale p est de $O(p\alpha^2 d_{\mathbf{G}} d_{\mathbf{H}} n_{\mathbf{G}} n_{\mathbf{H}})$. Autrement dit, notre noyau est calculable en temps polynomial.

4 Expériences

4.1 Classification supervisée

Les performances de nos noyaux ont été évaluées sur la base d’images **Core114** [2], dans le cadre de la classification multi-classes supervisée à l’aide de SVMs un-contre-un (cf. figure 7). Les noyaux entre histogrammes globaux de couleur nous sont apparus comme la méthode état de l’art la plus appropriée pour évaluer notre approche. Nous avons comparé les performances des noyaux d’appariement de sous-arbres virtuels avec ou sans pondération des régions (**wTW**, **TW**) avec les noyaux sur histogrammes globaux (**H**). Les paramètres des noyaux ont été ajustés par validation croisée au sein d’une boucle interne, les performances

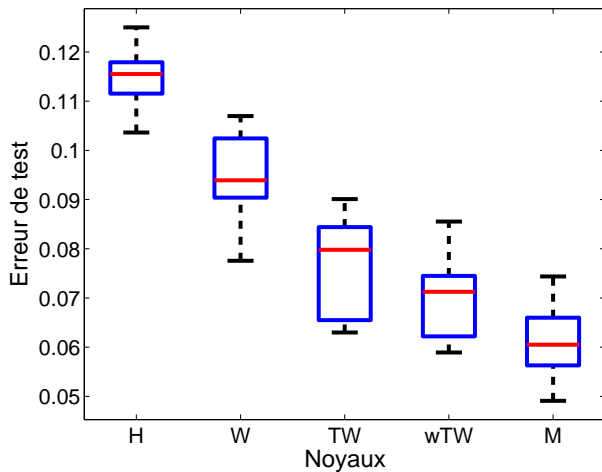


FIG. 7 – Performances en classification supervisée sur Corel14

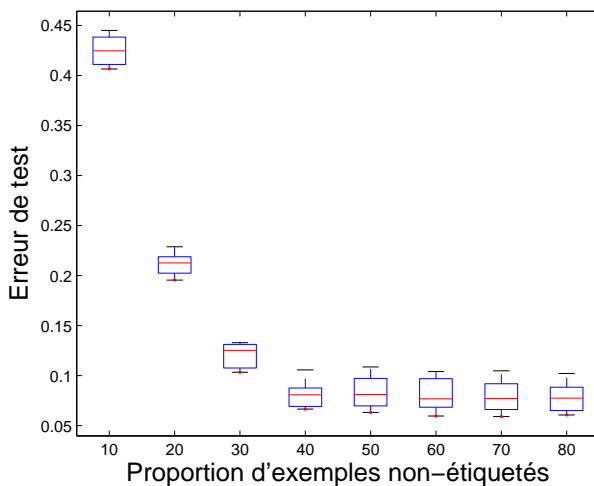


FIG. 8 – Performances en classification semi-supervisée sur Corel14

étant mesurées au sein d'une boucle externe. Nous présentons également les résultats (**M**) obtenus en apprenant à combiner suivant la méthode décrite dans [1] une centaine de noyaux pris avec différentes valeurs de α -arité, de profondeur de promenades arborescentes p , et de pondération d'aire de régions γ .

4.2 Classification semi-supervisée

Le contexte semi-supervisé permet de considérablement réduire le nombre d'exemples étiquetés nécessaires réaliser des performances satisfaisantes en classification. Par sa remarquable flexibilité, le cadre des méthodes à base de noyaux permet d'affronter cette tâche sans effort supplémentaire de développement d'algorithme. Nous avons ainsi utilisé nos noyaux avec LDS décrit dans [3]. Les diagrammes à boîtes dans la Figure 8 montrent que nos noyaux permettent d'atteindre les performances en apprentissage complètement supervisé pour seulement 10% des données étiquetées et 40% des données non-étiquetées.

5 Conclusion et perspectives

Nous avons proposé un cadre général pour les tâches de classification d'images, en nous appuyant sur des noyaux d'appariement de promenades arborescentes. Les noyaux utilisés avec des SVMs montrent des performances remarquables en classification supervisée sur des bases d'images naturelles. Des performances satisfaisantes ont également été observées en classification semi-supervisée. De nombreuses extensions de notre approche apparaissent naturellement. Des descripteurs locaux plus élaborés tels que les SIFT [8] peuvent être incorporés dans notre noyau de base. D'autres tâches, telles que l'indexation ou la détection peuvent aisément être envisagées dans le cadre des méthodes à base de noyaux.

Références

- [1] F. R. Bach, G. R. G. Lanckriet, and M. I. Jordan. Multiple kernel learning, conic duality, and the smo algorithm. In *Proc. ICML*, 2004.
- [2] O. Chapelle and P. Haffner. support vector machines for histogram-based classification. *IEEE Trans. Neural Networks*, 10(5) :1055–1064, 1999.
- [3] O. Chapelle and A. Zien. Semi-supervised classification by low density separation. In *Proc. AISTATS*, 2004.
- [4] R. Diestel. *Graph Theory*. Springer-Verlag, 2005.
- [5] D. Gusfield. *Algorithms on Strings, Trees, and Sequences*. Cambridge Univ. Press, 1997.
- [6] M. Hein and O. Bousquet. Hilbertian metrics and positive definite kernels on probability measures. In *AISTATS*, 2004.
- [7] F.-J. Huang and Y. LeCun. Large-scale learning with svm and convolutional nets for generic object categorization. In *Proc. CVPR*, 2006.
- [8] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *Int. J. Comp. Vision*, 60(2) :91–110, 2004.
- [9] P. Mahé, N. Ueda, T. Akutsu, J.-L. Perret, and J.-P. Vert. Extensions of marginalized graph kernels. In *ICML*, 2004.
- [10] F. Meyer. Hierarchies of partitions and morphological segmentation. In *Scale-Space and Morphology in Computer Vision*. Springer-Verlag, 2001.
- [11] J. Ramon and T. Gärtner. Expressivity versus efficiency of graph kernels. In *First International Workshop on Mining Graphs, Trees and Sequences*, 2003.
- [12] B. Schoelkopf, K. Tsuda, and J.-P. Vert. *Kernel methods in computational biology*. MIT Press, 2004.
- [13] J. Shawe-Taylor and N. Cristianini. *Kernel Methods for Pattern Analysis*. Cambridge Univ. Press, 2004.
- [14] F. Suard, V. Guigue, A. Rakotomamonjy, and A. Benschrair. Pedestrian detection using stereo-vision and graph kernels. In *IEEE Symposium on Intelligent Vehicule*, 2005.