

Apprentissage interactif par Online Gradient Boost en télédétection

Bertrand LE SAUX

ONERA - The French Aerospace Lab
F-91761 Palaiseau, France
bertrand.le_saux@onera.fr

Résumé – Cet article présente une approche pour l’apprentissage interactif en télédétection qui permet de définir des détecteurs *ad hoc* pour une mission et un contexte particuliers. Elle repose sur un algorithme d’Online Gradient Boost et introduit une nouvelle fonction de perte adaptée aux exemples fournis par un interprète. Elle est validée sur données réelles pour différents objets (bâtiments, végétation) ou contextes (imagerie optique ou radar) et pour plusieurs applications (détection de changements, transfert de détecteur à la caméra embarquée d’un drone).

Abstract – This paper presents an interactive learning approach for remote sensing that allows to build *ad hoc* detectors using a modified Online Gradient Boost. It is assessed on real data for various objects (buildings, vegetation) or contexts (optical or SAR images) and several applications (change detection, detector adaptation for use on on-board UAV camera).

1 Introduction

Les images satellitaires sont aujourd’hui disponibles en quantité tant pour les professionnels que pour le grand public. Les processus automatiques (segmentation, classification) ont longtemps permis d’indexer et d’exploiter ces données, cependant les résolutions de plus en plus fines offrent un niveau de détail qui permet d’extraire de plus en plus d’information de ces images. Cela conduit à une explosion du nombre de concepts visuels décelables dans une image. Il devient alors difficile de maintenir une base de détecteurs pour tous les types d’objets et toutes les situations. L’exploration interactive d’image est une voie prometteuse pour surmonter ce problème [9, 10, 11]. L’utilisateur indique ce qu’il recherche dans l’image, et le système cherche des régions similaires aux exemples fournis. Cela permet de s’adapter à des types d’images (capteur, résolution) et des contextes géographiques variés.

Plusieurs approches ont été proposées pour apprendre en ligne un concept visuel utile en télédétection. Dans [11], le système permettait de sélectionner des pixels représentatifs dans l’image pour estimer les distributions de colorimétrie de la requête (qui porte donc sur des pixels et non des objets). À l’inverse, des systèmes issus de la recherche d’image par le contenu comme VisiMine [4] ou [3] segmentaient l’image en de nombreuses imagettes et présentaient une liste ordonnée des résultats, ce qui conduit à perdre le contexte géographique. Les systèmes actuels incluent la dimension interactive par *active learning* [1, 12] basé sur des Machines à Vecteurs de Support (SVMs) mais génèrent une classification pixellique de l’image qui ne prend pas assez en compte la dimension objet.

Notre système combine les apports des précédentes approches. Il segmente l’image en patchs qui sont indexés pour représenter localement le voisinage d’une localisation. Pour conserver

l’information de contexte, l’interprète sélectionne les zones exemples avec une interface de type système d’information géographique. De plus, la sélection des exemples dans un système opérationnel pose deux problèmes. Des données d’apprentissage peuvent être mal-étiquetées (une zone mal tracée qui englobe des exemples de l’autre classe). Les ensembles d’apprentissage sont souvent déséquilibrés (il y a souvent plus d’exemples négatifs faciles à trouver que de positifs). Nous proposons une variante du Boosting qui permet l’apprentissage itératif en tenant compte de ces deux écueils.

L’apprentissage interactif par Online Gradient Boost avec une nouvelle fonction de perte est présenté en section 2. Cette approche est validée dans trois cas de figure : la détection de structures du sur-sol (section 3.1), l’adaptation de domaine pour la vision des drones (section 3.2) et la détection de changements (section 3.3) en imagerie satellite.

2 Apprentissage interactif

Interactivité et problématique Le processus d’apprentissage interactif est le suivant. L’interprète sélectionne des zones de l’image contenant l’objet de sa recherche et d’autres qui en sont exemptes (cf. Fig. 1-a). Des descripteurs x_k sont calculées sur des patchs P_k extraits uniformément de ces zones pour caractériser leur apparence visuelle. Associées au label y_k de la zone d’origine, ils constituent l’ensemble d’apprentissage $\mathcal{X} = \{(x_k, y_k)_{1 \leq k \leq N}, x_k \in \mathbb{R}^d, y_k \in \{-1, 1\}\}$. Un classifieur est entraîné sur cet ensemble puis appliqué aux patchs extraits de l’image entière (cf. Fig. 1-b). Le processus est ensuite itéré dans une boucle d’interactions avec l’utilisateur pour affiner le classifieur.

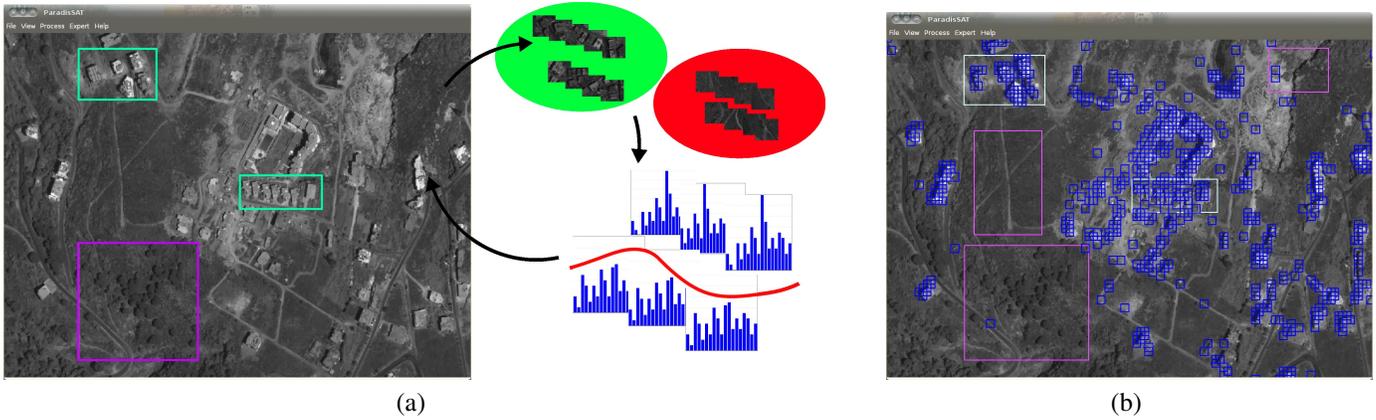


FIGURE 1 – (a) Apprentissage interactif par : 1. sélection de zones exemples de structures artificielles (rectangles verts) et de contre-exemples (rectangles mauves), 2. extraction de patches et 3. classification par Online Gradient Boost des Histogrammes d’Orientations de Gradients de chaque patch (b) Detections de structures artificielles sur des images QuickBird de résolution 0.6m après 2 itérations (carrés bleus)

Online Gradient Boost Les algorithmes de boosting cherchent à construire un méta-classifieur F à partir d’un ensemble de classifieurs faibles (ici les composantes de x_k). Online Gradient Boost est une variante incrémentale du boosting qui minimise le risque empirique :

$$R(F) = \frac{1}{N} \sum_k L(y_k, F(x_k)) \quad (1)$$

où L est une fonction de perte (par exemple $\exp(-yF(x))$ pour *adaboost*).

Nous avons montré qu’il est préférable pour notre application d’utiliser des fonctions non-convexes telles que la fonction DoomII [2] car le bruit d’étiquetage est particulièrement pénalisant pour les fonctions de perte convexes [6]. Pour gérer le problème des ensembles d’apprentissage déséquilibrés, les poids des erreurs de classification dans la minimisation itérative du risque empirique sont modifiés en fonction de la probabilité a priori $p(y)$ de chaque classe. La perte DoomII modifiée est alors définie par :

$$L(y, f(x)) = \frac{1 - \tanh(y \cdot f(x))}{p(y)} \quad (2)$$

3 Applications et résultats

3.1 Détection de structures artificielles

La classification de structures du sur-sol est utile tant en urbanisme (développement urbain) que pour la gestion de crise (par exemple pour la détection de camps de réfugiés suite à une catastrophe). Afin de discriminer ce type d’objets, les patches sont indexés par des Histogrammes d’Orientations de Gradient (HOG) qui offrent un excellent compromis performances / rapidité de calcul. Le classifieur construit par Online Gradient Boost sur ces descripteurs permet alors de détecter les bâtiments (cf. Fig 1-b).

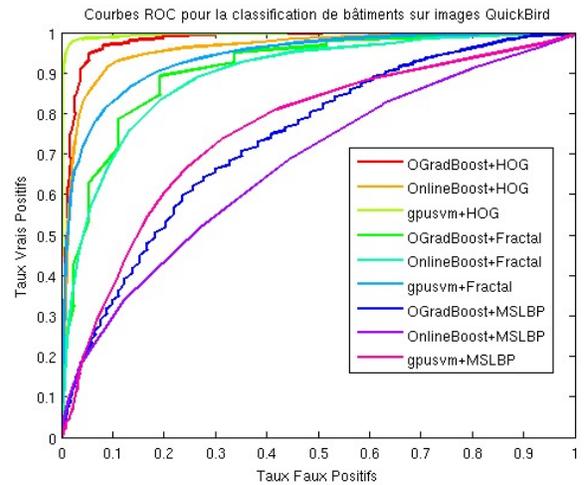
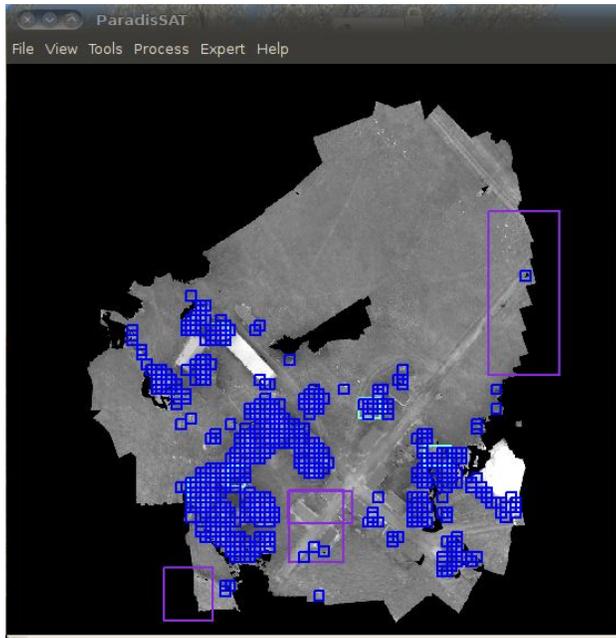
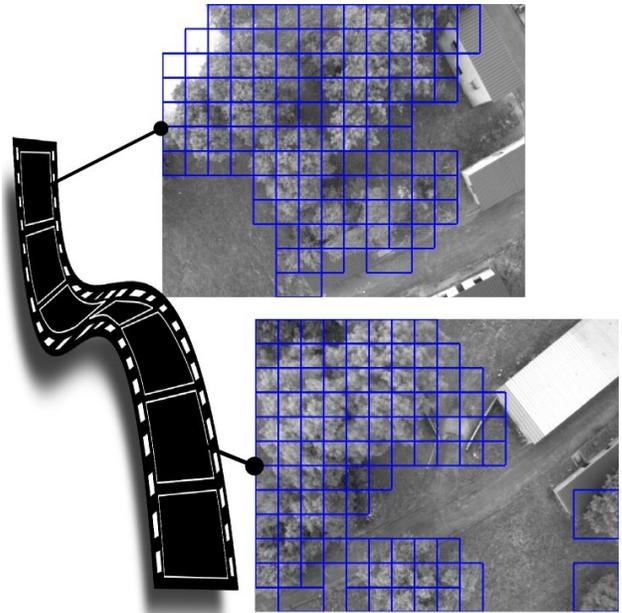


FIGURE 2 – Courbes ROC (*Receiver Operating Curve*) pour la classification de structures artificielles sur des images Quick-Bird de résolution 0.6m : 3 algorithmes Online Gradient Boost, Online Adaboost et SVM sont comparés avec 3 types de caractéristiques : HOG, Erreur Fractale et Motifs Binaires Multi-Échelle (MSLBP)

Les mesures de performance de classification (cf. Fig. 2) montrent que la combinaison Online Gradient Boost et HOGs obtient des résultats meilleurs que d’autres approches couramment utilisées en télédétection (différentes caractéristiques et différents classifieurs). Les résultats sont équivalents à une SVM non-incrémentale entraînée sur des HOGs pour un point de fonctionnement avec un taux de faux positifs $TFP > 10\%$, et ce avec une capacité de généralisation supérieure.



(a)



(b)

FIGURE 3 – (a) Détection dans une mosaïque de résolution 0.1m pour la cartographie (carrés bleus) (b) Adaptation géométrique du détecteur construit en ligne au domaine vidéo pour un usage embarqué (détectons = carrés bleus)

3.2 Adaptation de domaine pour la navigation des drones

L'apprentissage en ligne permet de s'adapter à des types d'objet spécifiques à un nouveau contexte. La figure 3 montre le design interactif d'un détecteur de végétation sur une orthomosaïque construite à partir d'une vidéo acquise par un drone. Deux objectifs sont visés.

En cartographie, les arbres sont des obstacles au vol du drone, notamment en phase d'atterrissage. Des détecteurs construits *ad hoc* permettent de cartographier la zone survolée et de planifier les vols suivants (cf. Fig. 3-a).

En détection embarquée, le classifieur est adapté au flux vidéo pour détecter des cibles ou des obstacles au moment où ils apparaissent dans le champ de vision du drone. Connaissant la position 3D et l'attitude du drone à un instant donné et les paramètres intrinsèques de la caméra, il est possible de calculer l'homographie H qui va permettre de recalculer l'image vidéo I_V dans la géométrie de l'ortho-mosaïque. Le classifieur issu du boosting est alors appliqué sur les patches :

$$P_k = g_H(m_k^{I_V}, I_V) \quad (3)$$

où $m_k^{I_V}$ sont des localisations réparties uniformément dans I_V et g_H une fonction de rectification d'image (cf. Fig. 3-b).

3.3 Détection de changements

La détection de changement en imagerie radar consiste à identifier les modifications (constructions nouvelles ou destructions) entre deux images acquises à deux dates différentes [8]. L'interprète désigne des zones avec changement du sur-sol et

des zones qui restent visuellement identiques. Si le processus d'apprentissage est identique aux cas précédents, les caractéristiques calculées mesurent la dissimilarité locale entre les deux images en deux étapes [5]. D'abord au niveau pixelique, le ratio de vraisemblance généralisé (*Generalized-Likelihood Ratio Test - GLRT* [7]) mesure la similarité entre les distributions des pixels des deux images. Dans un deuxième temps au niveau du patch, la distribution spatiale de la mesure GLRT est quantifiée par des HOGs. Cela permet de différencier les vrais changements du sur-sol par rapport à des régularités de la mesure dues au contexte géographique (orientation régulière des rues) ou à l'acquisition des données (erreur de recalage en raison d'angles de visée différents entre les deux images). La figure 4 montre que la méthode est capable de retrouver des constructions nouvelles y compris dans un milieu urbain dense.

4 Conclusion

Les différents domaines d'emploi explorés (image, vidéo à plus grande résolution, cartes de mesure de changement) montrent la généricité de notre approche pour construire des détecteurs d'objets ou de formes efficaces. Grâce au caractère intuitif des interactions, elle vise à transférer la conception de ces détecteurs du laboratoire à l'utilisateur final, l'interprète image, qui est le plus à même de définir l'objet de sa recherche.

Pour parvenir à un outil qui le satisfasse, les travaux futurs viseront à pouvoir distinguer des objets visuellement plus similaires, par exemple en utilisant des modèles structurés, et à rendre plus robuste l'adaptation des modèles de détection entre



FIGURE 4 – Détection de changement en milieu urbain par comparaison bi-date d’images TerraSAR-X. Image radar de gauche : des zones de construction nouvelles (rectangles pleins oranges) et de zones sans changements (rectangles pointillés verts) sont sélectionnées pour l’apprentissage, les changements détectés sont représentés par des carrés bleus. Colonne de droite : quelques exemples de comparaison bi-date dans le domaine optique montrent que les changements détectés correspondent à des modifications réelles du sur-sol (de haut en bas : destruction de bâtiments, installation de panneaux solaires, constructions nouvelles)

des images à des résolutions différentes et des vidéos.

Remerciements

Les auteurs remercient DigitalGlobe, Astrium Services et USGS pour avoir fourni des images pour cette étude, ainsi que le IEEE GRSS Data Fusion Technical Committee pour l’organisation du 2012 Data Fusion Contest.

Références

- [1] L. Bruzzone et C. Persello : Active learning for classification of remote sensing images. *In Proc. Int. Geosci. Remote Sens. Symp.*, Cape Town, South Africa, 2009.
- [2] N. Chauffert, J. Israël et B. Le Saux : Boosting for interactive man-made structure classification. *In Proc. Int. Geosci. Remote Sens. Symp.*, Munich, Germany, 2012.
- [3] M. Ferecatu et N. Boujemaa : Interactive remote-sensing image retrieval using active relevance feedback. *IEEE Trans. on Geosci. Remote Sens.*, 45(4):818–826, April 2007.
- [4] K. Koperski, G. Marchisio, S. Aksoy et S. Tusk : Visimine : interactive mining in image databases. *In Proc. Int. Geosci. Remote Sens. Symp.*, volume 3, pages 1810–1812, Toronto, Canada, 2002.
- [5] B. Le Saux et H. Randrianarivo : Urban change detection in sar images by interactive learning. *In Proc. Int. Geosci. Remote Sens. Symp.*, Melbourne, Australia, 2013.
- [6] C. Leistner, A. Saffari, P. Roth et H. Bischof : On robustness of on-line boosting - a competitive study. *In Proceedings of ICCV Workshop on On-line Learning for Comp. Vis.*, Kyoto, Japan, 2009.
- [7] P. Lombardo et C.J. Oliver : Maximum likelihood approach to the detection of changes between multitemporal SAR images. *IEE Proc. Radar, Sonar and Navig.*, 148(4):200–210, 2001.
- [8] G. Mercier, S. Derrode, E. Trouvé et L. Bombrun : Change detection in remote sensing observation. *In J. Chanussot, K. Chehdi et Ch Collet, éditeurs : Multivariate image processing : methods and applications.* Hermès, 2009.
- [9] M. Molinier, J. Laaksonen et T. Hame : Detecting man-made structures and changes in satellite images with a content-based information retrieval system built on self-organizing maps. *IEEE Trans. on Geosci. Remote Sens.*, 45(4):861–874, April 2007.
- [10] M. Quartulli et I. Olaizola : A review of EO image information mining. *ISPRS J Photogramm. Remote Sens.*, 75:11–28, 2013.
- [11] M. Schroeder, H. Rehrauer, K. Seidel et M. Datcu : Interactive learning and probabilistic retrieval in remote sensing image archives. *IEEE Trans. Geosci. Remote Sens.*, 38(5):2288–2298, May 2000.
- [12] D. Tuia, F. Ratle, F. Pacifici, M. Kanevski et W. Emery : Active learning methods for remote sensing image classification. *IEEE Trans. on Geosci. Remote Sens.*, 47(7): 2218–2232, 2009.