



HAL
open science

Classification de surfaces dans une image hyperspectrale urbaine par réseau de neurones pour la qualité de l'air

Valentin Boullisset, Jean-Luc Attié, Ronan Tournier, Nabil El Malki, Xavier Ceamanos-Garcia, Javier Andrey, Eric Pequignot

► To cite this version:

Valentin Boullisset, Jean-Luc Attié, Ronan Tournier, Nabil El Malki, Xavier Ceamanos-Garcia, et al.. Classification de surfaces dans une image hyperspectrale urbaine par réseau de neurones pour la qualité de l'air. Christelle Launois; Catherine Roussey. 9ème Conférence Nationale sur les Applications Pratiques de l'Intelligence Artificielle (APIA @ PFIA2023), Jul 2023, Strasbourg, France. AFIA-Association Française pour l'Intelligence Artificielle, Actes de la 9ème Conférence Nationale sur les Applications Pratiques de l'Intelligence Artificielle (APIA) - PFIA 2023, APIA (2023), pp.24-27, 2023. hal-04160817

HAL Id: hal-04160817

<https://hal.science/hal-04160817>

Submitted on 12 Jul 2023

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

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



Distributed under a Creative Commons Attribution 4.0 International License

Classification de surfaces dans une image hyperspectrale urbaine par réseau de neurones pour la qualité de l'air.

Valentin Boullisset¹, Jean-Luc Attié¹, Ronan Tournier², Nabil El Malki²,
Xavier Ceamanos-Garcia³, Javier Andrey⁴ and Eric Pequignot⁴

¹ LAERO, Université de Toulouse, Université Paul Sabatier, CNRS UMR5560, Toulouse, France

² IRIT, Université de Toulouse, Université Toulouse Capitole, CNRS UMR5505, Toulouse, France

³ CNRM, Météo-France, Université de Toulouse, CNRS UMR3589, Toulouse, France

⁴ WaltR-Parc Technologique du Canal, Ramonville Saint Agne, France

{valentin.boullisset, jean-luc.attie}@aero.obs-mip.fr

Résumé

Cet article est la première étape d'un processus complexe de mesure de la concentration de particules fines par des caméras spectrales. Pour ce faire, une classification précise et rapide de tous les pixels d'une image hyperspectrale est cruciale pour déterminer la nature des objets et leurs propriétés optiques. Les méthodes traditionnellement utilisées reposent sur la richesse de l'information spectrale avec des centaines de longueurs d'onde habituellement mesurées par les caméras. Cependant, négliger la dimension spatiale durant une classification conduit à certaines situations où les classes obtenues ne sont pas spatialement cohérentes. Nous proposons ici une méthode de classification optimisée pour la mesure de la qualité de l'air (Spatial Spectral Optimised Classification) basée sur l'apprentissage, qui prend en compte les caractéristiques spectrales mais aussi l'information spatiale via les pixels voisins.

Mots-clés

Classification, Image hyperspectrale, Apprentissage, Perception, SVM.

Abstract

This article is the first step of a complex process of measuring fine particle concentrations by spectral cameras. To do this, an accurate and fast classification of all pixels from a hyperspectral image is crucial to determine the nature of the objects and their optical properties. The methods traditionally used rely on the richness of the spectral information in measurements with hundreds of wavelengths. However, neglecting the spatial dimension during the classification can lead to situations where the resulting classes may not be spatially consistent. Here, we propose a classification method (Spatial Spectral Optimised Classification) based on learning, which takes into account spectral features but also spatial information via neighbouring pixels.

Keywords

Classification, Hyperspectral Imaging, Learning, Perception, SVM.

1 Introduction

La qualité de l'air est un enjeu de santé publique qui oblige les villes à la surveiller pour se conformer aux réglementations en vigueur depuis la loi LAURE¹ de 1996. Ces réglementations exigent une surveillance de la concentration de gaz et de particules dans l'air en raison de leur impact sur la santé. Les capteurs utilisés par les villes pour surveiller la qualité de l'air urbain sont limités en nombre et en portée, les modèles numériques compensent ces lacunes, mais ils ne sont pas assez contraints par les observations. L'installation de stations d'observation à grande échelle est trop coûteuse pour être une solution viable. L'objectif de notre travail est de permettre l'augmentation du nombre d'observations avec des solutions moins onéreuses, via, non pas des capteurs de mesure, mais une caméra spectrale placée au sol sur un point stratégique pour permettre une observation de la pollution à haute résolution et en temps réel sur une vaste zone urbaine à un coût raisonnable.

Pour identifier la pollution d'une ville à partir d'images, il faut d'abord associer les surfaces détectées à une valeur de réflectance de surface (capacité d'une surface à réfléchir l'énergie lumineuse incidente), qui est ensuite utilisée dans un processus de transfert radiatif et d'inversion pour déterminer les espèces chimiques présentes dans l'air et ainsi évaluer sa qualité. Pour ce faire, une image en RVB (RGB) ne suffit pas pour identifier avec précision des surfaces complexes dans une image, tandis qu'une image hyperspectrale, qui contient plusieurs dizaines de longueurs d'onde, est plus précise et permet d'obtenir des informations spécifiques. Les images hyperspectrales sont représentées comme un cube avec des dimensions spatiales et spectrales, et chaque pixel est associé à un vecteur qui permet d'obtenir une signature spectrale, ce qui permet de distinguer avec précision les différents objets présents dans l'image. L'article est organisé comme suit : la section suivante présente l'état de l'art ; la section 3 présente le matériel et les méthodes puis avant de conclure la section 4 présente les résultats.

1. Loi n°96-1236 du 30 décembre 1996 (NOR : ENVX9500163L)

2 État de l'art

Dans cet article nous nous concentrons sur la première étape du processus de mesure de la qualité de l'air qui consiste à identifier la nature de l'ensemble des pixels d'une image hyperspectrale. L'objectif consiste à attribuer à chaque pixel une classe prédéfinie qui correspond aux propriétés de la surface de l'objet. C'est pourquoi, nous optons pour une classification pixel par pixel qui permet de classifier l'ensemble des pixels d'une image. Les approches de détection d'objets tels qu'avec des CNN (Convolutional Neural Network) sont très efficaces pour détecter un ou plusieurs objets [1] mais ne permettent pas de classifier l'ensemble des pixels de l'image. Les méthodes de classification habituellement utilisées pour ce type de problématique comme SVM (support-vector machine) et les RF (random forest) traitent chaque pixel de manière indépendante en n'analysant que ses propriétés spectrales [2], sans considérer l'information sur les pixels voisins. Cependant, dans une image réelle, les pixels voisins sont corrélés. Négliger la notion de voisinage (dimension spatiale) en classification peut conduire à des situations où les classes obtenues ne sont pas cohérentes spatialement. Les profils morphologiques [3] ont été utilisés pour ce type de tâche mais le calcul pour chaque pixel d'une image hyperspectrale peut s'avérer très coûteux en termes de temps [4], ce qui pose un inconvénient majeur pour des applications nécessitant des résultats rapides en temps quasi réel sur des jeux de données volumineux.

Dans cet article, nous proposons une méthode de classification d'image hyperspectrale, optimisée pour la mesure de la qualité de l'air avec cohérence spatiale à l'aide d'un réseau de neurones et qui soit le plus rapide possible en termes de temps de calcul. La méthode commence par une réduction de dimension à l'aide d'une analyse en composante principale (ACP) pour limiter les problèmes liés au grand nombre de dimensions et réduire les temps de calcul. Puis les relations spatiales entre les pixels d'une image sont prises en compte grâce à une matrice de voisinage. Cette matrice est ensuite convertie en un vecteur et correspond à l'entrée d'un réseau de neurones.

3 Matériels, données et méthodes

3.1 Présentation des données

Les images sont acquises par une caméra hyperspectrale HySpex VNIR 1600² qui permet de faire des mesures sur 160 longueurs d'ondes de 400 à 1000 nm. Néanmoins, le but final est de capturer les images à partir de caméras bien moins onéreuses. Nous nous basons sur les caractéristiques de la future caméra WaltRCam constituée de 11 bandes spectrales et bien moins onéreuse que des caméras de type HySpex. Cet instrument, conçu par l'entreprise WaltR³, sera déployé fin 2023 pour couvrir la métropole Toulousaine. La caméra étant en phase finale de développement, des images ont été simulées en sélectionnant 11 lon-

gueurs d'onde de la caméra HySpex (414, 418, 422, 429, 461, 559, 609, 672, 782, 871, 941 nm) correspondant à la future WaltRCam. Les résultats présentés ici utilisent ces images simulées.

3.2 Classification hyperspectrale optimisée à une image d'une scène urbaine

Nous avons identifié les 6 classes les plus courantes qui représentent les surfaces visibles des objets de l'image. Plus tard dans la chaîne de traitement, ces classes seront utilisées pour calculer une valeur de réflectance à chaque pixel observé (permettant par la suite de déterminer la pollution de l'air). Les classes sont les immeubles éclairés (bâtiments éclairés directement par le soleil), les immeubles sombres (bâtiments non directement éclairés par le soleil), les toits, la végétation, le ciel et l'horizon (pixels éloignés plus difficilement identifiables). Il est à noter que la distinction entre immeuble éclairé et non éclairé peut être importante car les valeurs de réflectance associées à ces surfaces sont très différentes. Une partie des pixels de l'image est labellisée manuellement en fonction de la classe observée et sera nécessaire par la suite pour l'entraînement et le test des algorithmes d'apprentissage. Notre réseau de neurone optimisé en conséquence prend en entrée chaque pixel de l'image avec ses informations de voisinage et en sortie lui assigne une des 6 classes (avec un degré de confiance). La méthode de classification S₂OC (Spatial Spectral Optimised Classification, voir figure) que nous avons développée, basée sur l'apprentissage, se décompose en trois étapes :

- Réduction du nombre de dimensions via une analyse en composantes principales (ACP).
- Ajout de l'information spatiale via une matrice de voisinage pour chaque pixel.
- Classification de l'ensemble des pixels à l'aide du réseau de neurones.

Étape 1 : Réduction du nombre de dimensions à l'aide de l'analyse en composantes principales (ACP). Cette méthode linéaire pour la réduction de la dimensionnalité consiste à identifier les combinaisons linéaires maximisant la variance totale des données diminuant le nombre de dimensions. Cela permet de conserver l'information spatiale de l'image tout en condensant l'information spectrale qui réduit ainsi la dimension de l'image de 11 à 3 composantes avec une variance totale expliquée de près de 99%.

Étape 2 : Ajout de l'information spatiale avec une matrice de voisinage. Les méthodes de classification pour les images hyperspectrales actuellement utilisées se limitent souvent à attribuer à chaque pixel une classe basée uniquement sur ses propriétés spectrales, sans tenir compte de leur distribution spatiale des pixels. Pour pallier ce problème il est important de prendre en considération l'information spatiale avec la prise en compte du voisinage des pixels c'est à dire les pixels environnant proches. Une façon intuitive de définir le voisinage d'un pixel est de supposer que pour un pixel donné, ses voisins les plus proches appartiennent avec une forte probabilité à la même classe. Les zones homogènes présentes dans une image sont générale-

2. <https://www.hyspex.com>

3. <https://waltr.fr>

ment plus grandes que la taille d'un pixel. Par exemple, si un pixel représente la classe du ciel, il y a une forte probabilité que les pixels voisins soient également du ciel et appartiennent à la même classe.

Dans un premier temps chaque pixel de l'image est associé à une fenêtre de 7×7 pixels dont il est le centre. Cette fenêtre permet de construire la matrice de voisinage de taille $7 \times 7 \times 3$ correspondant au nombre de pixels de la fenêtre et des 3 composantes principales sélectionnées lors de la réduction de la dimension de l'étape précédente. Dans un second temps, pour chaque pixel, la matrice de voisinage, composée de $7 \times 7 = 49$ vecteurs de pixels à 3 dimensions est transformée en un vecteur de voisinage par dimension. Les vecteurs d'un pixel sont alors concaténés, formant ainsi un vecteur normalisé de $49 \times 3 = 147$ composantes. Chaque pixel de l'image possède son propre vecteur de voisinage qui regroupe les informations à la fois spectrales et spatiales de ce pixel. De cette façon, les dépendances spatiales sont prises en compte pendant la phase d'apprentissage automatique.

Étape 3 : Classification des pixels à l'aide de l'apprentissage. Cette troisième étape utilise un réseau de neurones pour assigner une classe à chaque pixel. La couche d'entrée correspond au vecteur normalisé de taille 147 contenant à la fois les informations spatiales et spectrales de l'image avec toutefois un nombre limité de composantes grâce à l'ACP. La fonction d'activation pour les couches cachées et d'entrée est une fonction sigmoïde. Cette fonction d'activation est généralement utilisée pour les problèmes de classification multi-classes car, donnant une valeur comprise entre 0 et 1, équivaut à une probabilité d'appartenance à une classe. Cette fonction est adaptée à la classification de type de surfaces qui est non linéaire. Afin de connaître la probabilité d'appartenance à une classe d'un pixel de l'image, les valeurs de la sortie du réseau doivent être normalisées pour obtenir des probabilités d'appartenance. La fonction softmax (fonction exponentielle normalisée) et utilisée comme fonction d'activation du vecteur de sortie. Cette fonction produit un vecteur de probabilité dans lequel la probabilité la plus élevée correspond à la classe prédite. La configuration optimale des poids synaptiques du réseau neuronal se fait par propagation arrière dans le but de minimiser la perte d'entropie croisée de la fonction d'erreur.

3.3 Métriques de performance

La performance du classifieur supervisé est évaluée de manière quantitative : les pixels préalablement étiquetés manuellement ont été comparés aux classes des pixels prédits par notre approche à l'aide d'une validation croisée afin d'augmenter la robustesse des résultats (et détecter un éventuel biais dans le jeu de données d'apprentissage). Les métriques de performance utilisées à cette fin sont des mesures classiques de rappel (recall), précision, F1-Score et d'exactitude (accuracy). Ces mesures servent à la fois à évaluer la méthode proposée mais aussi à la comparer par rapport à d'autres méthodes habituellement utilisées.

4 Résultats

4.1 Préparation de la base d'apprentissage

Notre jeu de données a été construit à l'aide de six classes distinctes et chaque pixel n'est associé qu'à une seule classe. La base d'apprentissage a été construite manuellement à l'aide de l'outil open source labelImg⁴. Des proportions similaires de pixels pour chaque classe ont été étiquetées pour limiter d'éventuelles erreurs ou biais lors de l'entraînement du réseau neuronal. Au total, près de 10000 pixels ont été étiquetés. Le jeu étiqueté de données est séparé en un ensemble d'entraînement (80% des pixels étiquetés) et un ensemble de test correspondant à 20%. Les ensembles sont modifiés aléatoirement pour permettre une validation croisée.

4.2 Comparaison avec les modèles SVM et SC

Dans cette partie, nous comparons les performances de la méthode S_2OC à SVM et SC (classification qui correspond à S_2OC sans le contexte de voisinage de chaque pixel).

Choix des hyperparamètres : Contrairement à SVM, les approches neuronales (SC et S_2OC) ont des hyperparamètres qui doivent être optimisés. Les trois hyperparamètres que nous avons cherchés à optimiser sont le nombre de couches cachées, le nombre de neurones par couche et le nombre d'itérations pour effectuer l'apprentissage. Notre stratégie consiste à tester diverses combinaisons d'architectures (ensembles d'hyperparamètres) et à sélectionner le réseau optimal qui minimise le mieux la fonction objectif. Nous avons testé plus d'une centaine d'architectures. Les hyperparamètres optimisés obtenus pour S_2OC sont 4 couches cachées de 350 neurones entraînés avec 100 itérations avec l'algorithme d'optimisation ADAM. L'entraînement d'un réseau de neurone n'étant pas déterministe (à paramètres et données équivalents, deux réseaux seront similaires mais différents), afin d'augmenter la robustesse des mesures d'évaluation, chaque modèle a été entraîné 5 fois en moyennant les résultats.

Résultats et comparaison : L'objectif est de déterminer les méthodes permettant d'obtenir les métriques optimales de performance sur notre base de test commune. Nous obtenons une accuracy 99,8% pour S_2OC suivi par SC et SVM avec 97,6% et 95,4%, respectivement. Les métriques précision, recall et F1-Score sont supérieures à 80%. Ceci est lié à la nature des données hyperspectrales plus précises que des images RVB et donc plus faciles à classifier. Quelle que soit la méthode, les classes végétation, horizon et ciel sont bien classifiées. Ceci est probablement dû au fait des caractéristiques spectrales qui sont nettes et avec peu de perturbations. Dans ce cas, précision, recall et F1-Score sont supérieurs à 99,6% montrant ainsi la facilité de classification. Les difficultés viennent avec les classes toit, immeuble éclairé et immeuble sombre. Pour cela l'approche S_2OC améliore les résultats par rapport à SC et SVM grâce à l'information spatiale du voisinage (un gain de performance

4. <https://pypi.org/project/labelImg/1.4.0/>

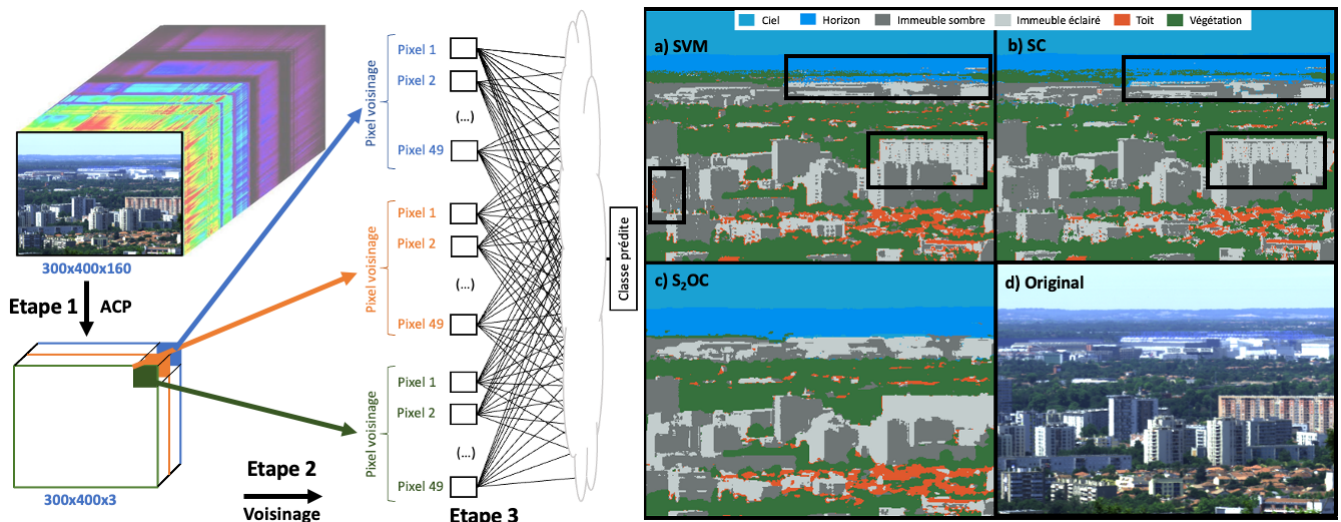


Figure : (Gauche) Schéma montrant la stratégie de classification spatiale et spectrale optimisée. L'étape 1 consiste à appliquer une analyse des composantes principales (ACP) sur l'image originale pour réduire sa dimension. L'étape 2 consiste à sélectionner la matrice de voisinage 7x7 pour inclure les caractéristiques spatiales et obtenir le vecteur de voisinage. L'étape 3 est l'application du réseau neuronal à l'aide de ce vecteur. (Droite) les résultats de classification de S₂OC, SC et SVM. Les rectangles représentent des exemples de zones d'incohérence.

allant jusqu'à 15% selon la classe et la métrique).

Pour compléter l'évaluation des modèles, une analyse visuelle est utile. La figure de droite montre les résultats de la classification. L'approche SVM a le plus d'incohérence d'un point de vue spatial (rectangles noirs sur les images) : observations isolées appartenant à des classes différentes de celles de leurs voisins (typiquement des pixels classés comme toit sur la façade des bâtiments). Les résultats de prédiction les plus cohérents sont obtenus par la méthode S₂OC. Il est également intéressant de noter que la frontière inférieure de la classe horizon pose de nombreuses difficultés aux classificateurs. L'introduction d'informations spatiales permet de lisser spatialement les résultats et de limiter les imprécisions de classification. Les frontières inférieures de la classe horizon en sont un bon exemple. Des résultats similaires ont été observés sur d'autres images montrant ainsi la bonne capacité de généralisation de l'approche S₂OC.

5 Conclusion

La première étape essentielle pour l'identification de la pollution d'une ville à partir d'images nécessite l'identification du type de surface des pixels de l'image pour pouvoir leur associer une valeur physique reflétant les caractéristiques de la surface de l'objet. La méthode S₂OC que nous avons développée, réduit d'abord le nombre de dimensions à traiter avant de prendre en compte les pixels adjacents grâce à une matrice de voisinage, puis utilise un réseau de neurones pour classifier chaque pixel. Cette approche est moins coûteuse en temps de calcul comparée aux méthodes de classification pixel par pixel actuellement utilisées. De plus, les résultats montrent que l'ajout de la matrice de voisinage améliore significativement les performances, avec des gains pouvant aller jusqu'à 15% par rapport aux méthodes SVM ou un réseau de neurones sans l'information de voisinage.

La suite des travaux consiste à associer une valeur de réflectance de surface à l'ensemble des pixels en fonction de la classe. Ceci afin de restituer l'épaisseur optique des aérosols, indice de pollution, à l'aide d'un modèle de transfert radiatif.

Remerciements

Ces travaux ont pu être effectués avec le soutien de la région Occitanie ainsi que le GIS neOCampus.

Références

- [1] N. Audebert, B. Saux, and S. Lefèvre, *Deep Learning for Classification of Hyperspectral Data : A Comparative Review*. *IEEE Geoscience and Remote Sensing Magazine*, **7** (2), pp. 159-173, 2019.
- [2] G. Lu, and B. Fei, *Medical hyperspectral imaging*. *J Biomed Opt*; **19** (1), pp. 523-565 2014.
- [3] M. Fauvel, Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. C. Tilton, *Advances in Spectral-Spatial Classification of Hyperspectral Images*. *Proceedings of the IEEE*; *Geosci*, **101** (3), pp. 652-675, 2013.
- [4] X. Huang and L. Zhang, *A comparative study of spatial approaches for urban mapping using hyperspectral ROSIS images over Pavia City, northern Italy*. *International Journal of Remote Sensing*; **30** (12), pp. 3205-3221, 2009.
- [5] K. Mounika, K. Aravind, M. Yamini, P. Navyasri, S. Dash and V. Suryanarayana, *Hyperspectral Image Classification using SVM with PCA*. *2021 6th International Conference on Signal Processing, Computing and Control (ISPCC)*; **107** (14), pp. 470-475, 2021.