

Segmentation non supervisée d’image hyperspectrale par mélange de gaussiennes spatialisé

Serge COHEN¹, Erwan LE PENNEC²

¹IPANEMA, Synchrotron Soleil
Saint-Aubin, France

²SELECT, INRIA Saclay IdF
Université Paris Sud, Orsay, France

`serge.cohen@synchrotron-soleil.fr, erwan.le_pennec@inria.fr`

Résumé – Cet article propose une nouvelle méthode de segmentation d’images hyperspectrales reposant sur un mélange de gaussiennes dont les proportions varient spatialement. Nous montrons comment estimer le nombre de classes ainsi que tous les autres paramètres ce modèle et en déduire une segmentation.

Abstract – This article introduces a new hyperspectral image segmentation method based on a Gaussian mixture whose proportions vary spatially. We show how to estimate the number of classes as well as all the other parameters and how to deduce a segmentation.

1 Introduction

La plateforme IPANEMA du synchrotron Soleil a pour objectif d’utiliser la *lumière* de Soleil pour étudier les matériaux anciens. Elle souhaite mettre à disposition de ses utilisateurs les instruments les plus sophistiquées à la fois pour les mesures physique et pour l’analyse des données. Elle offre notamment la possibilité d’acquérir des images hyperspectrales de haute résolution tant spatiale (de l’ordre de 100 000 pixels) que fréquentielle (de l’ordre de 1000 bandes de fréquences) ayant d’excellents rapports signal sur bruit. C’est à l’aide de cet outil qu’une étude sur les violons Stradivarius a été effectuée par Echard *et al.* [6]. Elle a permis de déterminer la technique utilisée par le célèbre facteur pour le vernis de ses violons. Ce travail a nécessité une étape fastidieuse de segmentation manuelle d’images hyperspectrales, étape que nous avons cherché à automatiser cette phase. L’objectif est donc d’obtenir une segmentation de ce type d’images en classes/régions homogènes sans connaître au départ ni la nature de ces classes ni même leurs nombres.

Cette problématique est classique en traitement des images hyperspectrales. Deux approches simples mais différentes sont naturelles : les approches spatiales (ou région) où les régions sont obtenus par des techniques de segmentation d’image regroupant localement des pixels dont les spectres sont similaires et et les approches spectrales (ou clustering) où les spectres sont groupés par classes similaires sans exploiter l’information spatiale. La première approche permet d’obtenir des composantes s’adaptant bien à la géométrie des images mais ne permet pas de reconnaître que deux zones disjointes correspondent au même type d’objet, la seconde ayant le comportement in-

verse. On peut citer, par exemple, pour la première école les travaux de Tarabalka *et al.* [11] ou Bunte *et al.* [4] et pour la seconde les travaux de Acito *et al.* [1] ou Yang *et al.* [12].

Il est alors naturelle de souhaiter combiner les avantages de ces deux approches. De nombreuses directions ont été proposées pour cela. On peut citer, par exemple, Tarabalka *et al.* [10] qui ont proposé de faire interagir ces deux méthodes en partant de la méthode spatiale pour déterminer des régions initiales fusionnées ensuite par des critères issues des méthodes spectrales ou les approches de type champs de Markov hiérarchique (cf Farag *et al.* [7]) dans lequel une régularisation spatiale est imposée sur les étiquettes issues du clustering.

Nous proposons ici une extension de la méthode spectrale permettant de prendre en compte, au moins partiellement, la nature géométrique des images. Elle repose sur des techniques d’estimation de densités conditionnelles et de sélection de modèles présentes en classification non supervisée et permet naturellement d’estimer simultanément un nombre de classes de spectres. La méthode la plus classique de classification non supervisée [3] repose sur une modélisation probabiliste des spectres par un mélange de gaussiennes ; une fois ce mélange estimé, un principe de maximum de vraisemblance permet d’attribuer à chaque spectre observé sa classe la plus probable et d’obtenir ainsi une segmentation de l’image. L’estimation du nombre de classes K se fait typiquement par des techniques de pénalisation [9]. Cette modélisation n’est cependant pas satisfaisante dans un cadre de segmentation : elle ne prend pas en compte le caractère spatial des données. Reprenant les idées de Kolaczyk *et al.* [8] et Antoniadis *et al.* [2], nous avons modélisé cette dépendance spatiale à travers les proportions du mélange : ceux-

ci varie en fonction de la position afin de prendre en compte l'inhomogénéité spatiale des images hyperspectrales. Notons que cette technique ne force pas une régularité sur les étiquettes mais uniquement sur la densité conditionnelle sous-jacente.

Dans la section 2, nous présenterons la méthode spectrale classique de la classification non supervisée par mélange de gaussiennes. Dans la section 3, nous montrerons comment prendre en compte le caractère spatial en étendant le modèle probabiliste précédent. Dans la section 4, nous décrirons plus précisément l'algorithme de minimisation utilisé. Enfin dans la section 5, nous présenterons une application à l'étude de vernis de violons Stradivarius.

2 Spectres, mélange de gaussiennes et classification non supervisée

L'imagerie hyperspectrale permet d'observer en chaque pixel x_i un spectre $Y_i \in \mathbb{R}^p$. L'objectif est alors d'assigner à chacun des pixels une classe à laquelle le spectre correspondant est estimé appartenir. Ceci suppose de déterminer ces classes ainsi que leur nombre.

Nous utilisons un modèle de type mélange de gaussiennes : la loi des spectres Y_i appartenant à la classe k est approchée par une gaussienne de moyenne μ_k et de matrice de covariance Σ_k dont la densité par rapport à la mesure de Lebesgue sera noté $\mathcal{N}_{\mu_k, \Sigma_k}$. Pour déterminer la loi de Y_i , il ne reste plus qu'à spécifier les probabilités π_k d'appartenir à la classe k . La densité de la loi de Y_i est donc supposé être proche de

$$\sum_{k=1}^K \pi_k \mathcal{N}_{\mu_k, \Sigma_k}(y).$$

Sous l'hypothèse que tous les spectres Y_i sont indépendants et de même loi, la loi de l'image observée est entièrement caractérisée par celle d'un pixel.

On peut alors chercher à estimer la loi commune de ces pixels par une des fonctions des collections de densités :

$$S(K) = \left\{ \sum_{k=1}^K \pi_k \mathcal{N}_{\mu_k, \Sigma_k}(y) \mid \sum_{k=1}^K \pi_k = 1, \mu_k \in \mathbb{R}^p, \Sigma \in [L, D, A]^K \right\}.$$

où $[L, D, A]^K$ impose une structure commune sur les matrices de covariances (sphérique, ellipse le long des axes ou libre, cf [3]) et K est un entier à déterminer.

Comme chez Acito *et al.* [1], à K fixé, les estimées $\hat{\pi}_k^{(K)}$, $\hat{\mu}_k^{(K)}$ et $\hat{\Sigma}_k^{(K)}$ s'obtiennent en minimisant la -log-vraisemblance

$$-\sum_{i=1}^N \log \left(\sum_{k=1}^K \pi_k \mathcal{N}_{\mu_k, \Sigma_k}(Y_i) \right).$$

Une règle de décision simple basée sur un principe de maximum de vraisemblance permet alors d'attribuer à chaque observation sa classe la plus vraisemblable :

$$\hat{k}[x_i] = \operatorname{argmax}_{1 \leq k \leq K} \hat{\pi}_k^{(K)} \mathcal{N}_{\hat{\mu}_k^{(K)}, \hat{\Sigma}_k^{(K)}}(Y_i).$$

Maugis et Michel [9] ont montré que l'on pouvait estimer un bon nombre de classe à partir des données et des estimés précédents en pénalisant la -vraisemblance par une quantité dépendant de K :

$$\hat{K} = \operatorname{argmin}_K - \sum_{i=1}^N \log \left(\sum_{k=1}^K \hat{\pi}_k^{(K)} \mathcal{N}_{\hat{\mu}_k^{(K)}, \hat{\Sigma}_k^{(K)}}(Y_i) \right) + \operatorname{pen}(K)$$

où $\operatorname{pen}(K)$ est une pénalité qui évite le sur-apprentissage si on la choisit de l'ordre de $\kappa \dim(S(K))$ avec κ assez grand. Leurs résultats garantissent que la densité du modèle à \hat{K} classes sélectionné est proche de la vraie densité des spectres. Bien que ce résultat n'implique pas de bons résultats de classification non supervisée d'un point de vue théorique, il s'avère que c'est le cas en pratique.

3 Segmentation non supervisée par mélange spatialisé

Le défaut principal de la méthode précédente est de ne pas prendre en compte la position des spectres : elle les considère comme des réalisations indépendantes de vecteurs aléatoires de même loi. Une manière simple de prendre en compte cette nature spatiale du problème est de faire varier la loi des spectres en fonction de la position. Comme souligné par Kolaczyk *et al.* [8], il est naturel de supposer que les lois des classes ne dépendent pas de la position et que la variation de la loi ne se fait donc qu'à travers les proportions du mélange.

En d'autres termes, on suppose maintenant que la densité par rapport à la mesure de Lebesgue de la loi de Y_i conditionnellement à la position du pixel x_i est proche de

$$\sum_{k=1}^K \pi_k(x_i) \mathcal{N}_{\mu_k, \Sigma_k}(y)$$

avec $\sum_{k=1}^K \pi_k(x_i) = 1$. C'est ce type de densité, ou plutôt de densité conditionnelle, que l'on appelle mélange de gaussiennes spatialisé.

On cherche alors à estimer la densité conditionnelle $s(y|x)$ de la loi des spectres par une fonction d'une des classes

$$S(K) = \left\{ \sum_{k=1}^K \pi_k(x) \mathcal{N}_{\mu_k, \Sigma_k}(y) \mid \forall x, \sum_{k=1}^K \pi_k(x) = 1, \mu_k \in \mathbb{R}^p, \Sigma \in [L, D, A]^K \right\}.$$

Cela revient à estimer un nombre de classes K , des moyennes μ_k , des matrices de covariances Σ_k ainsi que une fonction de mélange $\pi(x)$. En cherchant à estimer une fonction, on rentre dans le monde des statistiques non paramétriques, une forme de régularisation est donc nécessaire. On propose ici de suivre encore Kolaczyk *et al.* [8] et d'utiliser des proportions qui sont constantes par morceaux sur une partition de l'image. Soit \mathcal{P} une telle partition, on note ses régions \mathcal{R}_l et, pour un nombre fixé K de classes, on considère ainsi les densités condition-

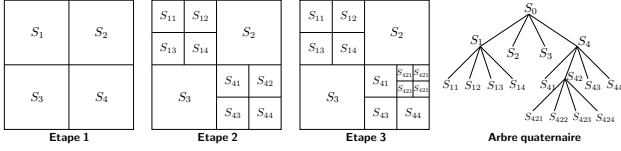


FIGURE 1 – Partition récursive et arbre quaternaire

nelles appartenant à l'ensemble

$$S(K, \mathcal{P}) = \left\{ \sum_{k=1}^K \left(\sum_{\mathcal{R}_l \in \mathcal{P}} \pi_k[\mathcal{R}_l] \mathbf{1}_{x_i \in \mathcal{R}_l} \right) \mathcal{N}_{\mu_k, \Sigma_k}(y) \mid \forall \mathcal{R}_l \in \mathcal{P}, \sum_{k=1}^K \pi_k[\mathcal{R}_l] = 1, \mu_k \in \mathbb{R}^p, \Sigma \in [L, D, A]^K \right\}$$

Il s'agit alors d'estimer un nombre de classes K , des moyennes μ_k , des matrices de covariances Σ_k , une partition \mathcal{P} ainsi des proportions de mélange $\pi[\mathcal{R}_l]$ sur chacune des régions de la partition.

Dans un article sur l'estimation de densités conditionnelles[5], nous avons montré que l'on peut estimer simultanément tous ces paramètres si l'on force la partition à appartenir à une classe structurée, par exemple la classe des partitions récursives en carrés dyadiques illustrées Figure 1. À nombre de classes K et partition \mathcal{P} fixés, un principe de maximum de vraisemblance (implémentable grâce à l'algorithme EM) permet d'estimer des moyennes $\hat{\mu}_k^{(K, \mathcal{P})}$, des matrices de covariances $\hat{\Sigma}_k^{(K, \mathcal{P})}$ ainsi que les proportions $\hat{\pi}_k^{(K, \mathcal{P})}[\mathcal{R}_l]$. Le choix d'un nombre de classes \hat{K} et d'une partition $\hat{\mathcal{P}}$ s'effectue alors en minimisant une -log vraisemblance pénalisée

$$\sum_{i=1}^N -\log \left(\sum_{k=1}^K \left(\sum_{\mathcal{R}_l \in \mathcal{P}} \hat{\pi}_k^{(K, \mathcal{P})}[\mathcal{R}_l] \mathbf{1}_{x_i \in \mathcal{R}_l} \right) \mathcal{N}_{\hat{\mu}_k^{(K, \mathcal{P})}, \hat{\Sigma}_k^{(K, \mathcal{P})}}(Y_i) \right) + \text{pen}(K, \mathcal{P})$$

où $\text{pen}(K, \mathcal{P})$ est encore de la forme $\kappa \dim(S(K, \mathcal{P}))$. Ce choix est justifié théoriquement et permet de bien estimer la densité conditionnelle. Il ne reste plus alors qu'à appliquer le principe de maximum de vraisemblance de la section précédente pour attribuer des classes aux spectres observés :

$$\hat{k}_i = \underset{1 \leq k \leq \hat{K}}{\text{argmax}} \left(\sum_{\mathcal{R}_l \in \hat{\mathcal{P}}} \hat{\pi}_k^{(\hat{K}, \hat{\mathcal{P}})}[\mathcal{R}_l] \mathbf{1}_{x_i \in \mathcal{R}_l} \right) \mathcal{N}_{\hat{\mu}_k^{(\hat{K}, \hat{\mathcal{P}})}, \hat{\Sigma}_k^{(\hat{K}, \hat{\mathcal{P}})}}(Y_i).$$

4 Algorithme

L'estimation de ces paramètres ne peut être réalisée en pratique par une exploration exhaustive sur les différentes partitions, leur nombre augmentant de manière exponentielle avec la taille de l'image. En combinant les principes des fonctions majorantes de l'algorithme EM et des techniques de programmation dynamique utilisant la structure des partitions dynamiques, nous avons implémenté (dans le logiciel MIXMOD[3]) un algorithme itératif effectuant une estimation simultanée, à K fixé, de tous les autres paramètres.

Le point de départ est celui de l'algorithme EM où l'on remarque que si l'on pose

$$P_k^{i, (n)} = P(k_i = k | x_i, y_i, \mathcal{P}^{(n)}, \pi^{(n)}, \mu^{(n)}, \Sigma^{(n)}) \propto \left(\sum_{\mathcal{R}_l \in \mathcal{P}^{(n)}} \pi_k^{(n)}[\mathcal{R}_l] \mathbf{1}_{x_i \in \mathcal{R}_l} \right) \mathcal{N}_{\hat{\mu}_k^{(n)}, \hat{\Sigma}_k^{(n)}}(Y_i)$$

alors on peut majorer la -log vraisemblance par une quantité plus simple séparable en (\mathcal{P}, π) et (μ, Σ) :

$$\begin{aligned} L(\mathcal{P}, \pi, \mu, \Sigma) &= \sum_{i=1}^N -\log \left(\sum_{k=1}^K \left(\sum_{\mathcal{R}_l \in \mathcal{P}} \hat{\pi}_k^{(K, \mathcal{P})}[\mathcal{R}_l] \mathbf{1}_{x_i \in \mathcal{R}_l} \right) \mathcal{N}_{\mu_k^{(K, \mathcal{P})}, \Sigma_k}(Y_i) \right) \\ &\leq \sum_{i=1}^N \sum_{k=1}^K -P_k^{i, (n)} \log \left(\sum_{\mathcal{R}_l \in \mathcal{P}} \pi_k[\mathcal{R}_l] \mathbf{1}_{x_i \in \mathcal{R}_l} \right) \\ &\quad + \sum_{i=1}^N \sum_{k=1}^K -P_k^{i, (n)} \log(\mathcal{N}_{\mu_k, \Sigma_k}(Y_i)) + \text{Cst}^{(n)} \end{aligned}$$

avec égalité en $(\mathcal{P}^{(n)}, \pi^{(n)}, \mu^{(n)}, \Sigma^{(n)})$. On utilise alors la structure de la pénalité

$$\text{pen}(K, \mathcal{P}) = \kappa (|\mathcal{P}|(K-1) + Kp + \dim([L, D, A]^K))$$

ainsi que la structure de la partition pour obtenir

$$\begin{aligned} L(\mathcal{P}, \pi, \mu, \Sigma) + \text{pen}(K, \mathcal{P}) &\leq \sum_{\mathcal{R}_l \in \mathcal{P}} \left(\sum_{k=1}^K - \left(\sum_{x_i \in \mathcal{R}_l} P_k^{i, (n)} \right) \log(\pi_k[\mathcal{R}_l]) + \kappa(K-1) \right) \\ &\quad + \sum_{i=1}^N \sum_{k=1}^K -P_k^{i, (n)} \log(\mathcal{N}_{\mu_k, \Sigma_k}(Y_i)) + \text{Cst}'^{(n)} \end{aligned}$$

avec égalité en $(\mathcal{P}^{(n)}, \pi^{(n)}, \mu^{(n)}, \Sigma^{(n)})$. Si $(\mathcal{P}^{(n+1)}, \pi^{(n+1)}, \mu^{(n+1)}, \Sigma^{(n+1)})$ minimise la fonctionnelle de droite alors

$$\begin{aligned} L(\mathcal{P}^{(n+1)}, \pi^{(n+1)}, \mu^{(n+1)}, \Sigma^{(n+1)}) + \text{pen}(K, \mathcal{P}^{(n+1)}) \\ \leq L(\mathcal{P}^{(n)}, \pi^{(n)}, \mu^{(n)}, \Sigma^{(n)}) + \text{pen}(K, \mathcal{P}^{(n)}) \end{aligned}$$

avec une inégalité stricte dès lors que la valeur de la fonctionnelle de droite a diminué de manière stricte par rapport à sa valeur en $(\mathcal{P}^{(n)}, \pi^{(n)}, \mu^{(n)}, \Sigma^{(n)})$. La minimisation de cette fonctionnelle en (μ, Σ) s'obtient à l'aide d'une formule close (la même que celle de l'algorithme EM classique) tandis qu'un algorithme de type CART exploitant la structure additive sur les régions permet de minimiser de manière rapide la partie en (\mathcal{P}, π) : pour tout carré dyadique \mathcal{R} , une formule close donne la proportion $\pi^{(n+1)}[\mathcal{R}]$ optimale pour chaque carré dyadique tandis qu'un algorithme récursif, basé sur l'idée que le meilleur partition d'un carré dyadique est soit ce carré lui-même soit l'union des meilleurs partitions de ses quatre sous-carrés, optimise efficacement la partition. Il en résulte un algorithme qui diminue efficacement la -log vraisemblance à chaque étape et est utilisable pour des dimensions de p de l'ordre de la centaine.

5 Applications

Nous avons utilisé cet algorithme sur les données déjà étudiées par Echard *et al.* [6]. La figure 5 présente deux segmentation distinctes : une (à gauche) obtenue par la méthode purement spectrale de la section 2 et une (à droite) obtenue par la méthode décrite dans les sections 3 et 4 qui prend en compte la nature spatiale des données. Le paramètre κ apparaissant dans la pénalisation a été estimé par une méthode de type *heuristique de pente*[9] dans les deux cas.

Les résultats obtenus sont similaire : le nombre de classes estimé est le même et les paramètres des gaussiennes sont proches. La segmentation obtenue dans le cadre spatiale (visible sur la figure) permet cependant de diminuer fortement le nombre de points aberrants de la segmentation purement spectrale : le modèle de proportions constant par morceaux favorise bien les classes selon leur proportion dans la région proche et non pas selon leur proportion dans l'image entière. Cette régularisation sur les étiquettes n'est cependant pas explicite comme dans les modèles markoviens. Les segmentations obtenues ont été jugées pertinentes par les utilisateurs.

Cette méthodologie a été testée sur d'autres jeux de données réelles avec le même succès. Il manque cependant une phase de validation systématique sur des données synthétiques ainsi qu'une comparaison avec les méthodes de l'état de l'art (en particulier avec les méthodes spatiales et les méthodes mixtes). Il nous faut notamment comprendre si cette prise en compte du caractère spatiale dans les proportions permet une régularisation suffisante (y compris dans des cas plus bruités que les spectres acquis au synchrotron Soleil) ou si la régularité sur les étiquettes (ou leurs transitions) doit être imposée explicitement.

Le résultat présenté ici correspond ainsi plus à une première étude de faisabilité de la méthode qu'à un résultat définitif. Outre celle de la régularisation, de nombreuses questions restent encore ouverte. Nous avons, par exemple, proposé une calibration raisonnable pour l'estimation de densité de la pénalité par heuristique de pente et la question de son efficacité pour la segmentation non supervisée est ouverte. Bien réduire de dimension, une nécessité en pratique, est également crucial : nous avons utilisé une réduction de dimension par simple ACP mais nous travaillons également sur des techniques moins linéaires, de type K-SVD. L'objectif reste de fournir la solution automatisée la plus efficace possible aux utilisateurs d'IPANEMA.

Références

- [1] N. ACITO, G. CORSINI et M. DIANI : An unsupervised algorithm for hyperspectral image segmentation based on the gaussian mixture model. *In Proc. IGARSS*, volume 6, pages 3745–3747, 2003.
- [2] A. ANTONIADIS, J. BIGOT et R. von SACHS : A multiscale approach for statistical characterization of functional images. *J. Comput. Graph. Statist.*, 18(1):216–237, 2008.
- [3] Ch. BIERNACKI, G. CELEUX, G. GOVAERT et F. LANGROGNET : Model-based cluster and discriminant analy-

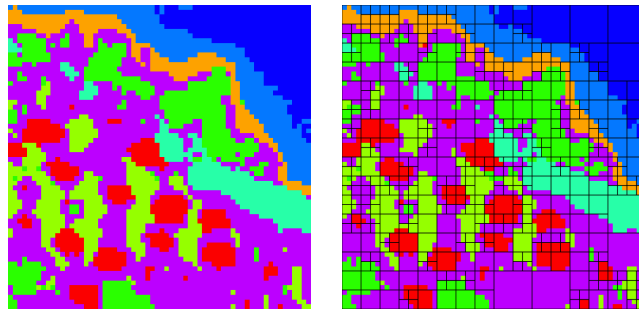


FIGURE 2 – Classification non supervisée et segmentation non supervisée (avec la partition optimale)

- sis with the MIXMOD software. *Comput. Statist. Data Anal.*, 51(2):587–600, 2006.
- [4] M. BUNTE, D. R. THOMPSON, R. CASTANO, S. CHIEN et R. GREELEY : Metric learning for hyperspectral image segmentation. *In IEEE WHISPERS*, 2011.
- [5] S. COHEN et E. LE PENNEC : Conditional density estimation by penalized likelihood model selection. Submitted, 2011.
- [6] J.-Ph. ECHARD, L. BERTRAND, A. von BOHLEN, A.-S. LE HÔ, C. PARIS, L. BELLOT-GURLET, B. SOULIER, A. LATTUATI-DERIEUX, S. THAO, L. ROBINET, B. LAVÉDRINE et S. VAIEDELICH : The nature of the extraordinary finish of Stradivari's instruments. *Angewandte Chemie International Edition*, 49(1):197–201, 2010. ISSN 1521-3773.
- [7] A. FARAG, R. MOHAMED et A. EL-BAZ : A unified framework for map estimation in remote sensing image segmentation. *IEEE Trans. Geosci. Remote Sens.*, 43(7):1617–1634, 2005.
- [8] E. KOLACZYK, J. JU et S. GOPAL : Multiscale, multi-granular statistical image segmentation. *J. Amer. Statist. Assoc.*, 100(472):1358–1369, 2005.
- [9] C. MAUGIS et B. MICHEL : A non asymptotic penalized criterion for Gaussian mixture model selection. *ESAIM : P & S*, 2009. To appear.
- [10] Y. TARABALKA, J. A. BENEDIKTSSON, J. CHANUSSOT et J. C. TILTON : Multiple spectral-spatial classification approach for hyperspectral data. *IEEE Trans. on Geoscience and Remote Sensing*, 48(11):4122–4132, 2010.
- [11] Y. TARABALKA, J. CHANUSSOT et J. BENEDIKTSSON : Segmentation and classification of hyperspectral data using watershed transformation. *Pattern Recognition*, 43(7):2367–2379, 2010.
- [12] J.-M. YANG, P.-T. YU et B.-C. KUO : A nonparametric feature extraction and its application to nearest neighbor classification for hyperspectral image data. *IEEE Trans. Geosci. Remote Sens.*, 48(3):1279–1293, 2010.