

Imagerie par faisceau d'ions : application à l'étude des objets du patrimoine

Astrid TAZZIOLI¹ Alexandre GIRARD² Quentin LEMASSON² Laurent PICHON² Claudine LOISEL³ François MIRAMBET² Claire PACHECO²

¹Centre de Recherche et de Restauration des Musées de France, 14 quai François Mitterrand, 75001 Paris, France

²EDF R&D Lab, 6 Quai Watier, 78400 Chatou, France

³Laboratoire de Recherche des Monuments Historiques, 29 rue de Paris, 77420 Champs-sur-Marne, France

Résumé – L'imagerie par faisceau d'ions permet d'étudier les objets du patrimoine en révélant notamment leur état d'altération, leur provenance, leurs techniques de fabrication. Souvent hétérogènes et stratifiés, ces objets nécessitent une analyse 2D de leur surface pour obtenir une vision représentative du matériau. Les techniques employées déterminent leur composition chimique et leur stratigraphie, mais génèrent rapidement des milliers de spectres à interpréter, rendant leur traitement manuel très chronophage. Cette étude propose une façon d'accélérer le traitement de ces cubes de données, en regroupant les voxels issus de zones homogènes, qui peuvent alors être traités simultanément. Pour cela, les résultats de différents algorithmes de clustering et de segmentation seront comparés. L'application de la méthodologie porte sur des grisailles, peintures à base de plomb fréquemment rencontrées sur les vitraux depuis le Moyen-Age.

Abstract – Ion beam imaging enables the study of heritage objects by revealing their state of degradation, provenance and manufacturing techniques. These objects are often heterogeneous and layered, requiring 2D analysis for a representative view. The techniques used determine their chemical composition and layering but rapidly generates thousands of spectra that must be interpreted, making manual processing highly time-consuming. This study aims to accelerate the analysis of these data cubes by grouping voxels from homogeneous areas. These spectra could be simultaneously processed. To reach that purpose, different clustering and segmentation algorithms are compared. The application of the methodology focuses on grisailles, lead paintings, commonly found in stained-glass windows.

1 Imagerie par faisceau d'ions pour l'étude des objets du patrimoine

1.1 Objets du patrimoine

Les objets du patrimoine couvrent une vaste gamme d'artefacts, chacun se distinguant par ses matériaux, sa période et ses techniques de fabrication. Ils incluent des peintures, des sculptures, des textiles ainsi que divers objets artisanaux. La céramique et la verrerie, à travers la poterie, les mosaïques ou les vitraux, illustrent un savoir-faire technique spécifique. Certains de ces objets mesurent à peine quelques centimètres, tandis que d'autres, bien plus imposants et de toutes les formes, s'intègrent à des édifices entiers.

Ces objets sont hétérogènes, connaître leur composition chimique informe sur les processus d'altération, leur provenance mais aussi les techniques de fabrication alors utilisées. Ces objets sont aussi stratifiés, connaître la composition chimique globale seulement est insuffisante pour l'interprétation et nécessite d'étudier la répartition des éléments chimiques en profondeur. Des analyses non invasives et non destructives sont privilégiées dans le domaine du patrimoine, comme les analyses par faisceau d'ions (IBA pour *Ion Beam Analysis*). Visualiser la répartition spatiale des éléments chimiques, comme le plomb, a notamment permis de distinguer le verre, le lavis et les grisailles sur un panneau de la Cathédrale de Chartres [12], déterminer quantitativement les épaisseurs de peinture rendrait compte aussi des techniques de fabrication.

1.2 Analyse par faisceau d'ions

Les IBA reposent sur les interactions entre les ions incidents et la matière de la cible à analyser. Deux d'entre elles sont utilisées ici : l'analyse PIXE (*Particle Induced X-Ray Emission* [11]) et l'analyse RBS (*Rutherford Backscattering Spectroscopy* [16]). Sous l'effet du faisceau, l'analyse PIXE mesure les rayons X émis par les atomes de la cible, chaque pic correspondant à une transition électronique d'énergie caractéristique d'un élément chimique et d'intensité proportionnelle à sa concentration chimique. Le PIXE basse énergie (BE) permet de quantifier les éléments majeurs constitutifs de la matrice alors que le PIXE haute énergie (HE) se focalise sur les éléments traces en atténuant le signal des éléments majeurs. Néanmoins, le PIXE offre une faible résolution en profondeur. L'analyse RBS, quant à elle, mesure l'énergie et le nombre de particules rétrodiffusées après collision avec les atomes de la cible. Elle permet ainsi de reconstruire la stratigraphie de la cible en fonction de l'énergie des particules rétrodiffusées, reliée à leur profondeur. Elle offre une bonne résolution en profondeur mais distingue difficilement des éléments chimiques de masse proche. Ces deux techniques sont complémentaires car le traitement des spectres RBS sans les résultats PIXE reste ambigu. En effet, le RBS repose sur un problème inverse, où plusieurs compositions peuvent générer le même spectre. [9].

1.3 Objectif et présentation de la méthodologie

Jusqu'à présent, les images 2D-PIXE/RBS étaient traitées de manière qualitative (en comparant des intensités des pics) et non quantitative (en déterminant des concentrations chimiques en ppm et des épaisseurs en nm). Cette étude vise à établir une méthodologie pour le traitement quantitatif du cube de données spectrales issues de différentes techniques afin de reconstruire la stratigraphie fine d'objets du patrimoine. L'approche proposée ici repose sur le regroupement des voxels en zones localement homogènes, afin de diminuer le nombre de spectres à traiter, tout en soulignant les différences entre les régions. Le regroupement peut être effectué soit sur les données spectrales seulement, soit sur les caractéristiques spatiales et spectrales. Les algorithmes d'apprentissage non supervisé ont été préférés à ceux supervisés car ces derniers requièrent un ensemble de données préalablement traitées et étiquetées, ce qui n'est pas réalisable sans effectuer un prélèvement d'une certaine surface et pour la grande variété des objets analysés à l'accélérateur.



FIGURE 1 : Méthodologie globale pour la reconstruction de la stratigraphie fine d'objets du patrimoine

Cette méthodologie (Figure 1) a été développée d'abord sur des échantillons de référence simples, constitués de quelques éléments chimiques, parfaitement connus et fabriqués en laboratoire puis appliquée à des matériaux plus complexes, similaires à ceux rencontrés dans le domaine du patrimoine, notamment les grisailles.

La méthodologie présentée à la Figure 1 est détaillée ici, étape par étape, et illustrée sur les volumes spectraux acquis sur une grisaille. Cet échantillon a été fabriqué par une restauratrice de façon à créer des zones d'intérêt, correspondant à différentes épaisseurs de grisaille. Le taux de comptage global, en tant que somme des intensités sur tous les canaux, contient les informations les plus caractéristiques du cube de données spectrales. L'image de ce taux sera alors considérée comme une vérité-terrain.

2 Pré-traitement des données PIXE et RBS

2.1 Combinaison des données spectrales

La première étape de la méthodologie consiste à combiner des données spectrales : pour chaque pixel du cube, un nouveau spectre est artificiellement formé à partir de la concaténation des voxels correspondant aux spectres PIXE et RBS normalisés (Figure 2). Cette concaténation s'effectue directement au niveau des canaux d'acquisition des détecteurs, car les signaux issus de ces deux techniques reposent sur des phénomènes physiques différents et ne sont donc pas directement comparables en énergie. En effet, les canaux du PIXE correspondent à des énergies de photons tandis que ceux de la RBS représentent des énergies de particules rétrodiffusées. L'objectif est ainsi de construire un spectre composite intégrant les informations

spécifiques de chaque technique, sans chercher à comparer directement leurs énergies. Le cube formé contient à la fois

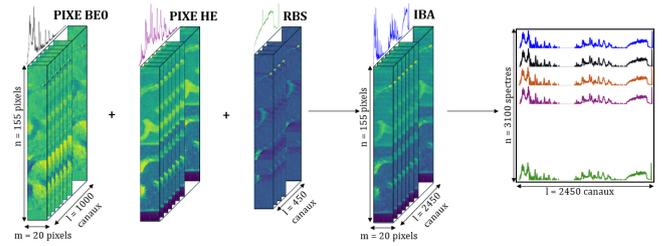


FIGURE 2 : Concaténation des données spectrales issues des techniques PIXE et RBS pour créer artificiellement de nouveaux spectres contenant à la fois les informations de composition (PIXE) et de stratigraphie (RBS).

l'ensemble des informations de composition chimique élémentaire mais aussi de stratigraphie de la zone analysée. Les régions issues du regroupement des voxels sont homogènes en composition et ont des stratigraphies communes. Cependant ces cubes de données sont composés de milliers de spectres, eux-mêmes codés sur des milliers de canaux. Une première étape de réduction des données est incontournable pour éviter la malédiction de la dimensionnalité [3] qui conduit à l'attribution d'aucun cluster par les algorithmes de partition des données.

2.2 Réduction des données spectrales

La réduction des données spectrales est légitime car de nombreuses bandes à l'intérieur des spectres sont corrélées. Des méthodes telles que l'analyse par composantes principales (ACP) ou la factorisation par des matrices non négatives ont été testées sur nos données [14]. Néanmoins, elles ne sont pas adaptées à notre problématique car les données issues de l'analyse RBS sont non linéaires [9]. À l'inverse, l'algorithme UMAP (*Uniform Manifold Approximation and Projection* [10]) garantit à la fois la préservation de la structure globale et locale des données lors de la réduction. Le cube de données réduites, à 2 dimensions avant le clustering ou à 3 dimensions avant la segmentation, représente fidèlement le cube de données dans l'espace original. L'objectif de UMAP n'est pas d'approximer les données au sens de l'ACP mais de conserver les relations entre celles-ci. Le choix des valeurs d'hyperparamètres de UMAP a été effectué en sélectionnant la combinaison de valeurs qui maximise la corrélation des distances [8] dans l'espace latent par rapport à l'original pour préserver la structure globale des données, ainsi que la fidélité [15] pour s'assurer de la préservation de la structure locale. UMAP s'appuie sur une représentation des données sous forme de graphes de voisinage à la fois dans l'espace initial et dans l'espace latent. L'objectif va alors être de préserver au maximum la correspondance entre ces deux graphes de voisinage tout en garantissant la structure de corrélation (ce qui est obtenu à partir d'une entropie croisée entre les graphes des deux espaces). La fidélité va correspondre à une mesure globale sur le graphe de voisinage de l'espace latent pour déterminer si les points proches dans l'espace latent correspondent à des points proches dans l'espace initial. Il existe une mesure réciproque, la discontinuité, pour déterminer si les points proches dans

l'espace original le sont aussi dans l'espace latent.

3 Regroupement des pixels

Après réduction des données, le regroupement des pixels en régions est alors facilité. Il peut être effectué en se basant seulement sur les spectres (à travers le clustering) ou en utilisant les caractéristiques spatiales et spectrales des données (à travers la segmentation). De nombreux algorithmes existent, pour lesquels un certain nombre d'hyperparamètres, propres à chacun, doivent être sélectionnés attentivement.

3.1 Clustering

Des algorithmes de clustering sont appliqués sur les données réduites en 2D. Des méthodes basées sur la densité comme DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*) [7] ou sa version améliorée avec un ajout de critère hiérarchique HDBSCAN (*Hierarchical Density-Based Spatial Clustering of Applications with Noise*) [5] sont particulièrement adaptées car UMAP repose sur des critères de densité. HDBSCAN est privilégiée car elle est très rapide et la sélection des hyperparamètres plus facile. L'homogénéité intra-cluster et la séparation inter-classe sont évaluées pour déterminer la combinaison d'hyperparamètres représentant au mieux nos données réduites.

3.2 Segmentation

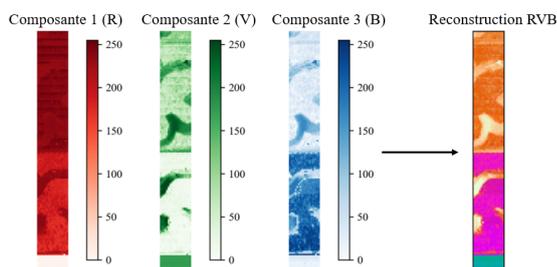


FIGURE 3 : Données réduites par UMAP à 3 dimensions (avec métrique euclidienne, 15 voisins et distance minimale de 0) puis reconstruction de l'image RVB avant la segmentation.

Sur les données réduites en 3D, une image RGB est reconstruite à partir de ces trois composantes (Figure 3) et des algorithmes tels que Quickshift ont été testés. Quickshift [17] dérive de l'algorithme Mean Shift [6] en introduisant une hiérarchie entre pixels voisins. Un arbre hiérarchique reliant les pixels en fonction de leur densité et de leur similarité est construit. Le choix des hyperparamètres a été optimisé en effectuant une recherche par grille des différentes combinaisons et en sélectionnant celle minimisant les Critères d'Information d'Akaike (AIC) [1] et Bayésien (BIC) [13]. Le premier privilégie une partition avec la meilleure vraisemblance alors que le deuxième tient compte à la fois de cette vraisemblance mais aussi du nombre de clusters.

3.3 Comparaison du clustering et de la segmentation

Les partitions obtenues après clustering et segmentation (Figure 4), correspondent aux zones demandées pour la fabrication des éprouvettes de grisaille, ressemble à l'image du taux de comptage du cube de données et sont proches les unes des autres.

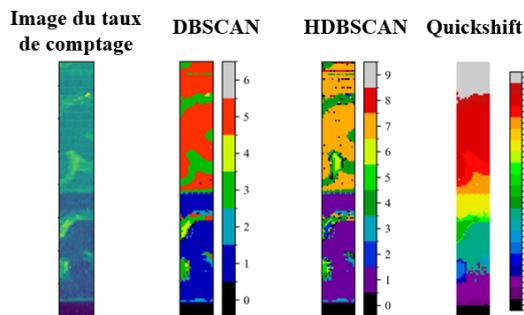


FIGURE 4 : Partition des données selon l'algorithme utilisé et comparaison au taux de comptage des données PIXE et RBS concaténées.

DBSCAN (métrique euclidienne, 30 voisins, epsilon = 0.471)
HDBSCAN (métrique euclidienne, 30 voisins, alpha = 1)
Quickshift (voisinage = 5, taille des régions = 10, ratio = 0.5)

D'après la Table 1, Quickshift est la méthode qui minimise à la fois les critères de log-vraisemblance à travers l'AIC et le BIC. La segmentation par Quickshift est alors privilégiée.

TABLE 1 : Scores de clustering et de segmentation. Quickshift est l'algorithme minimisant ces deux critères.

Algorithme	nombre de clusters	AIC	BIC
DBSCAN	8	62902	62944
HDBSCAN	10	62901	62962
Quickshift	21	60840	60967

Les clusters formés sont représentatifs des différentes zones, comme l'indiquent les histogrammes des résidus entre l'ensemble des spectres d'un cluster et son estimateur qui est le spectre médian. Les partitions ne sont jamais parfaites, notamment aux frontières entre clusters. L'objectif est alors d'obtenir les meilleurs regroupements qui résument nos données en un certain nombre de clusters sans pour autant lisser excessivement les différences spectrales.

4 Traitement des spectres regroupés et reconstruction de la stratigraphie associée

Une fois les pixels regroupés en clusters, l'analyse a été ramenée à une liste de spectres 1D. Les spectres médians, canal par canal, c'est-à-dire la concaténation marginale des médians par bande en première approche, de chacune de ces régions sont

extraits à la fois pour les techniques PIXE et RBS. En effet, à la différence des spectres sommes et des spectres moyens, la médiane est résistante aux anomalies et aux pixels mal attribués. L'avantage principal est l'obtention de spectres médians bien moins bruités que les spectres originaux du cube de données spectrales. La limite inhérente est quant à elle le lissage des différences intra-cluster.

Les spectres médians PIXE sont alors traités par GUPIX, logiciel classique de traitement PIXE ([4]). Celui-ci repose sur des calculs basés sur la théorie des interactions particules-matière, selon les conditions expérimentales, en modélisant les spectres PIXE générés par chacun des éléments chimiques présents et en corrigeant les effets de matrice. La composition obtenue est donc utilisée pour interpréter les spectres médians RBS. Ces derniers sont traités grâce au logiciel DataFurnace [2] : des spectres correspondant à une stratigraphie sont simulés et ajustés itérativement jusqu'à convergence.

A partir de la stratigraphie déterminée, une image de la stratigraphie de la grisaille peut être reconstruite, comme représentée à la Figure 5.

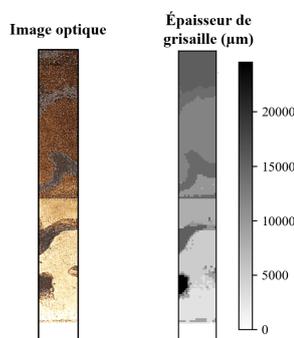


FIGURE 5 : Image optique et représentation spatiale des épaisseurs de grisaille.

La grisaille appliquée est composée d'oxydes de plomb, de fer, de zinc et de manganèse. La stratigraphie est en escalier, avec, comme attendu de bas en haut, le verre et ensuite une simple, puis une double couche de grisaille. Les épaisseurs sont respectivement évaluées à 4.9 et 11.9 μm . Sur chacune de ces couches, un dessin forme une surépaisseur, soit une épaisseur totale de 14.1 à 24.6 μm de grisaille.

5 Conclusion et perspectives

Un large panel de techniques de réduction des données, de méthodes de clustering et de segmentation non supervisées a été testé. La segmentation, en tenant compte des caractéristiques à la fois spatiales et spectrales des données réduites fournit une partition plus fidèle du matériau que le clustering.

Selon la question posée sur l'objet, si l'objectif est d'étudier un certain élément chimique dans le volume spectral, il est possible de restreindre les données aux canaux d'intérêt de cet élément et ainsi d'orienter la partition en fonction de l'objectif de l'analyse.

L'outil développé et présenté ici permet de reconstruire la stratigraphie de la zone analysée, avec une représentation plus visuelle, et intelligible par tous les profils. Il évite le traitement quantitatif de chacun des spectres du cube en traitant

seulement les spectres médians des zones homogènes. La réponse obtenue est alors rapide et de meilleure qualité car les spectres médians sont bien moins bruités que les originaux. Leur traitement, en particulier RBS, est donc facilité. Le principal inconvénient, inhérent à la méthodologie est de lisser les différences d'épaisseur et de concentration à l'intérieur d'un cluster. Si ce lissage est un frein pour l'utilisation de cette méthodologie, il est encore tout à fait possible de traiter l'ensemble du volume spectral voxel par voxel mais au détriment de la résolution spectrale et de la durée de calcul.

Remerciements

Nous remercions chaleureusement l'Atelier MurAnèse, atelier spécialisé dans la restauration de vitraux patrimoniaux, pour avoir fabriqué ces éprouvettes selon nos souhaits.

6 Bibliographie

Références

- [1] H. AKAIKE : A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723, décembre 1974. Conference Name : IEEE Transactions on Automatic Control.
- [2] N.P. BARRADAS et C. JEYNES : Advanced physics and algorithms in the IBA DataFurnace. *Nuclear Instruments and Methods in Physics Research Section B : Beam Interactions with Materials and Atoms*, 266(8):1875–1879, avril 2008.
- [3] R. BELLMAN : Curse of dimensionality. *Adaptive control processes : a guided tour*. Princeton, NJ, 3(2), 1961.
- [4] J.L. CAMPBELL, D.J.T. CUREATZ, E.L. FLANNIGAN, C.M. HEIRWEGH, J.A. MAXWELL, J.L. RUSSELL et S.M. TAYLOR : The Guelph PIXE Software Package V. *Nuclear Instruments and Methods in Physics Research Section B : Beam Interactions with Materials and Atoms*, 499:77–88, juillet 2021.
- [5] R. JGB CAMPELLO, D. MOULAVI et J. SANDER : Density-based clustering based on hierarchical density estimates. *In Pacific-Asia conference on knowledge discovery and data mining*. Springer, 2013.
- [6] D. COMANICIU et P. MEER : Mean shift : a robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5):603–619, mai 2002.
- [7] M. ESTER, H.-P. KRIEGEL et X. XU : A density-based algorithm for discovering clusters in large spatial databases with noise. *In Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, 1996.
- [8] A. GRACIA, S. GONZÁLEZ, V. ROBLES et E. MENASALVAS : A methodology to compare Dimensionality Reduction algorithms in terms of loss of quality. *Information Sciences*, 270:1–27, juin 2014.

- [9] M. M. LI, W. GUO, B. VERMA, K. TICKLE et J. O'CONNOR : Intelligent methods for solving inverse problems of backscattering spectra with noise : a comparison between neural networks and simulated annealing. *Neural Computing and Applications*, 18(5):423–430, juin 2009.
- [10] L. MCINNIS, J. HEALY et J. MELVILLE : UMAP : Uniform Manifold Approximation and Projection for Dimension Reduction, 2020.
- [11] P. MORETTO et L. BECK : Émission X induite par particules chargées (PIXE) : théorie. *Techniques d'analyse*, décembre 2003.
- [12] C. PACHECO, Q. LEMASSON, B. MOIGNARD, L. PICHON, M. RADEPONT et D. GOURIER : D'AGLAE à New AGLAE. *Technè*, 2016.
- [13] G. SCHWARZ : Estimating the Dimension of a Model. *The Annals of Statistics*, 6(2):461 – 464, 1978.
- [14] Astrid TAZZIOLI, Quentin LEMASSON, Alexandre GIRARD, Laurent PICHON, Brice MOIGNARD et Claire PACHECO : Advanced 2D-PIXE/RBS processing with Machine Learning at the New AGLAE facility for ancient layered objects. *Nuclear Instruments and Methods in Physics Research Section B : Beam Interactions with Materials and Atoms*, 555:165469, octobre 2024.
- [15] M. C. THRUN, J. MÄRTE et Q. STIER : Analyzing Quality Measurements for Dimensionality Reduction. *Machine Learning and Knowledge Extraction*, 5(3):1076–1118, août 2023.
- [16] P. TROCELLIER et P. TROUSLARD : Spectrométrie de collisions élastiques et de réactions nucléaires. Théorie. *Techniques d'analyse*, juin 2002.
- [17] A. VEDALDI et S. SOATTO : Quick Shift and Kernel Methods for Mode Seeking. In *Computer Vision – ECCV 2008*, Berlin, Heidelberg, 2008. Springer Berlin Heidelberg.