Détection et classification de décors gravés sur des céramiques anciennes par analyse d'images

Teddy Debroutelle

Centre Val de Loire UNIVERSITÉ Spirit of the Loire Valley

UNIVERSITÉ D'ORLÉANS

UNIVERSITE D'ORLEANS

ÉCOLE DOCTORALE DE MATHÉMATIQUES, INFORMATIQUE, PHYSIQUE THÉORIQUE ET INGÉNIERIE DES SYSTÈMES

LABORATOIRE : PRISME

Thèse présentée par :

Teddy DEBROUTELLE

soutenue le : **19 Février 2018**

pour obtenir le grade de : **Docteur de l'Université d'Orléans**

Discipline/ Spécialité : **Traitement des images et du signal**

Détection et classification de décors gravés sur des céramiques anciennes par analyse d'images

Thèse dirigée par :

Sylvie TREUILLET	MCF HDR , Université d'Orléans
Aladine CHETOUANI	MCF, Université d'Orléans
Matthieu EXBRAYAT	MCF, Université d'Orléans

RAPPORTEURS :

Olivier LEZORAY	PRU, Université de Caen
Frédéric MORAIN-NICOLIER	PRU, Université de Reims

JURY :

Jenny BENOIS-PINEAU	PRU, Université de Bordeaux 1
Jean-Yves RAMEL	PRU, Université de Tours
Marcilio PEREIRA DE SOUTO	PRU, Université d'Orléans

Table des matières

Liste des tableaux

Table des figures

Chapitre 1

Introduction

1.1 Projet ARCADIA

En 1968, les travaux d'aménagement d'un lac artificiel sur le site de La Médecinerie situé à Saran dans le département du Loiret ont mis au jour les vestiges d'un ensemble d'ateliers de production céramique (voir Fig. 1.1). Une première campagne de fouilles dirigée par Jean Chapelot de 1969 à 1972 permit de découvrir une dizaine de fours à proximité des restes de l'ancienne voie romaine Orléans-Chartres. Au milieu des années 1990, Sébastien Jesset, alors à l'Institut National de Recherches Archéologiques Préventives (INRAP), a mené une recherche sur le matériel céramique issu des premières fouilles et sa diffusion régionale en collaboration avec la Fédération Archéologique du Loiret (FAL) [11].

FIGURE 1.1 – Site archéologique "La Médecinerie".

La datation de ce site couvre la période Mérovingienne et Carolingienne du haut Moyen Âge (VIe -IXe siècle). C'est un site unique en France, voire en Europe dans cette période chronologique tant par le nombre de structures de production recensées que par son organisation. L'établissement de ce centre s'explique par la proximité et l'abondance des ressources naturelles nécessaires à l'élaboration des

terres cuites : l'argile, le sable et l'eau mais aussi le bois indispensable à la cuisson, qui était très certainement prélevé dans la Forêt d'Orléans. La proximité des voies de transports permettait la diffusion de la production.

En l'absence de source textuelle mentionnant les ateliers de Saran, la compréhension de ce patrimoine, de l'histoire économique et de celle des techniques repose sur le travail de l'archéologue. En dehors du terrain et de la fouille des structures, l'archéologue pratique des études sur la céramique elle-même. Ces études s'appuient sur des observations macroscopiques (morphologie, étude des décors, etc.), microscopiques (études pétrographiques) voir physicochimiques (analyses élémentaires) et entendent in fine définir la fonction, préciser la datation et identifier la provenance de l'objet.

Sur le site de La Médecinerie, il apparaît que la production est essentiellement une céramique commune destinée à une utilisation domestique.
La plupart des tessons font apparaître un décor imprimé dans l'argile fraîche par le potier sur le pourtour de la panse ou de l'épaule de la céramique [12]. Ce procédé de décoration dit « à la molette » n'est pas spécifique aux ateliers de Saran, ni à la période du Haut Moyen Âge. Il est plutôt répandu durant l'antiquité tardive [13]. La molette est un petit cylindre tournant sur un axe fixé à un manche (sorte de roulette large) que le potier aura pris soin de sculpter de façon personnelle (voir Fig. 1.2). Néanmoins, peu d'exemplaires de cet outil sont connus, car la molette était souvent réalisée dans un matériau périssable : entièrement en bois ou manche en bois et molette en os, au mieux en métal.

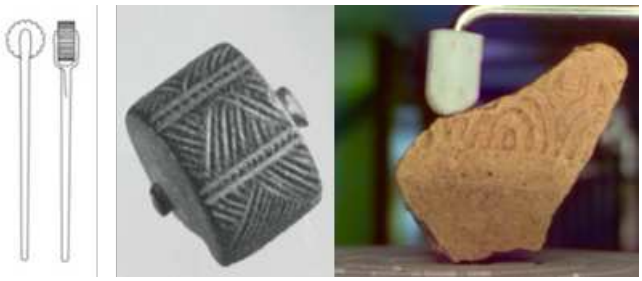


FIGURE 1.2 – Exemple de molette servant à graver un décor sur de la céramique.

Comme aucun exemplaire n'a été extraits lors des fouilles, il y a de fortes présomptions pour que les molettes utilisées à Saran aient été en bois [12]. Compte-tenu de l'usure relativement rapide d'une molette en bois, les céramiques marquées d'une même molette ont forcément été produites par un même potier dans un temps relativement court. Le motif « à la molette » s'avère donc un marqueur chronologique très important pour l'archéologue. L'analyse stylistique des décors des tessons, comme signatures des potiers, jumelée au contexte stratigraphique et à leur relation à des structures bien datées (datation absolue des fours par archéomagnétisme) permettent à l'archéologue de préciser la chronologie des répertoires céramiques et de mieux comprendre l'organisation des ateliers et la diffusion de la production vers les centres de consommation.

Le site de La Médecinerie, dont la fouille n'avait pas été menée à son terme dans les années 70, a été ré-ouvert à partir de 2008 dans le cadre d'une étude menée par Sébastien Jesset [14]. Cette première étude portant sur environ mille tessons s'est appuyée sur un relevé d'empreinte manuel et une appréciation visuelle pour distinguer environ 104 molettes ou matrices de référence différentes mais de motifs assez peu diversifié : pour l'essentiel, il s'agit de frises avec des motifs en chevrons, carrés, triangles, ou losanges.

Le corpus dégagé lors des dernières fouilles depuis 2009 atteint aujourd'hui près de 38000 tessons avec ou sans décor. Face à l'augmentation du corpus, la méthode manuelle n'est plus possible. Une numérisation des tessons et une reconnaissance automatique des décors deviennent donc indispensables. C'est l'objectif du projet ARCADIA (Automatic Recognition of Ceramics Achieved by Digital Image Analysis) financé par la région Centre. En mettant en commun les compétences en céramiques archéologiques de la FAL, celles d'analyse d'images du Laboratoire PRISME, et apprentissage automatique du Laboratoire Informatique d'Orléans, ce projet vise à développer un outil générique qui réalise automatiquement l'association entre différents décors et leurs matrices pour faciliter l'interprétation et la mise en valeur d'un patrimoine céramique archéologique. Un tel outil de classification automatique n'existe pas à ce jour dans le domaine de l'archéologie. Comme le classement d'artefacts archéologiques se fonde sur la répétition des observations,

on comprend bien l'intérêt que peut revêtir l'analyse numérique d'images ou d'empreintes 3D pour les décors pour l'analyse du patrimoine céramique en général.

1.2 La céramique et la vision par ordinateur

La céramique est peut-être le matériel archéologique le plus étudié en Vision. Plusieurs travaux ont été publiés visant à automatiser la reconstitution d'un vase complet à partir des tessons retrouvés ou au classement des céramiques : soit du point de vue de la forme (morphologie de la céramique), soit de la technologie (composition-cuisson de la pâte), ou encore des décors. La thèse de Maiza [15] dresse un état de l'art de tous les projets antérieurs à 2005 qui se sont intéressés à la reconnaissance des formes de céramiques. Bien que ces projets apportent des contributions différentes, ils mettent en œuvre des méthodes d'acquisition et des modèles communs. Comme il s'agit d'analyser une forme 3D, ils font appel à un procédé d'acquisition tridimensionnel pour numériser les artefacts et modélisent la céramique comme une surface de révolution [16, 17], dont ils cherchent à estimer les paramètres par le biais de différentes approches, parfois inspirées des solutions techniques utilisées par les archéologues [16]. Cette modélisation par un axe de symétrie (ou axe de rotation) et une courbe plane (méridienne ou profil pour les archéologues) est directement liée au processus de façonnage du vase sur un tour de potier. L'axe et la courbe de profil sont ensuite utilisés pour automatiser le relevé jusqu'alors réalisé à la main par l'archéologue [16], ou pour apparier de façon automatique des tessons et des formes de référence [15], ou encore proposer une solution de remontage automatique [17]. Des publications récentes montrent qu'il s'agit encore d'un problème ouvert [18, 19, 20] avec de nouvelles propositions pour déterminer le couple axe/profil et de nouvelles méthodes de remontage automatique.

Cependant, ces études restent limitées par ce modèle idéal de surface de révolution qui n'est pas forcément généralisable à tous les types de tessons céramiques. En effet, seules les céramiques tournées (fabriquées au tour) peuvent être considérées comme axialement symétriques, et même dans ce cas, les conditions de

production artisanale font que l'axe n'est pas le même sur toute la hauteur du pot. D'autre part, des tessons voire des céramiques entières répondent d'avantage à la géométrie de la sphère, caractérisée par un couple centre-rayon, pour lesquels les méthodes de détermination axe-profil échouent immanquablement. Les caractéristiques morphologiques n'ont pas été les seules à faire l'objet d'un traitement numérique. Un plus petit nombre d'études se sont penchées sur la caractérisation d'attributs technologiques des céramiques. Les couleurs et la texture des surfaces intérieures et extérieures des tessons, ainsi que de la tranche, sont les éléments étudiés par les archéologues pour déterminer le type d'argile utilisé et également le type de cuisson. Classiquement, l'archéologue décrit la teinte de la pâte, en s'aidant parfois d'un nuancier spécialisé. Cependant sans mesure physique, la description de la couleur, même avec un nuancier, reste subjective et sujette aux conditions d'illumination ambiantes. Des études se sont intéressées à une méthode de calibrage colorimétrique du dispositif de prise de vue, adapté aux céramiques [21] pour obtenir une correspondance précise entre une photographie numérique des pâtes de céramiques et le nuancier de Munsell, sous différentes conditions d'illumination [22]. D'autres travaux couplent une acquisition tridimensionnelle avec une acquisition infrarouge et hyperspectrale [23]. L'objectif est alors d'évaluer la pertinence de l'acquisition hyperspectrale comme procédé non destructif de caractérisation minéralogique des pâtes et de les comparer avec des données acquises sur le terrain. Ces articles restent toutefois très méthodologiques et se bornent à démontrer le potentiel de ces outils de mesures physiques sans résultats tangibles d'un point de vue typologique et classification des céramiques. Les travaux qui proposent une réelle approche de classement automatique du point de vue de la technologie de fabrication évacuent la problématique de mesure de la couleur en prenant comme point de départ une acquisition photographique dans un environnement contrôlé (illuminant constant et utilisation d'une mire de correction). Deux études abordent ce problème de classification des tessons céramiques par la texture du matériau [24, 25], en exploitant des caractéristiques fréquentielles des images extraites à l'aide de filtres (ou ondelettes) de Gabor [26]. En se fondant sur une référence établie par des archéologues, la performance de cette caractérisation de texture est évaluée par des méthodes d'apprentissage automatique : par un clustering flou [24] ou une recherche des plus proches voisins [25]. Smith et al.

[27] proposent quant à eux de coupler texture et couleur. La mise en correspondance entre tessons et base de données se fait sur deux niveaux par un système de votes. En premier lieu, ils recherchent les dix points d'intérêt les plus similaires (plus proches voisins) entre une image de référence et les autres en s'appuyant sur un descripteur de texture local inspiré de SIFT [2]. Ensuite, une mesure de similarité est appliquée pour sélectionner le meilleur des dix points d'intérêt en comparant des histogrammes locaux de couleur par la distance du χ^2. De façon plus complète, Makridis et Daras [28] extraient sur les deux faces des tessons un ensemble de caractéristiques locales (contours, texture, couleur), puis les quantifient par clustering (K-means) avec une sélection des clusters les plus discriminants par Analyse en Composantes Principales (ACP). Le classement automatique est ensuite réalisé par une recherche des K plus proches voisins (K-ppv) dans l'espace multidimensionnel du descripteur. Les auteurs montrent que leur descripteur est plus performant sur leur corpus que des descripteurs habituellement utilisés dans des problématiques plus génériques d'indexation d'images. Si les résultats de ces études sont généralement prometteurs, il convient de rester critique vis-à-vis de leur échantillonnage car, dans la plupart des cas, il est impossible d'apprécier la variabilité du corpus étudié. De plus, il y a parfois un mélange entre caractérisation de la pâte elle-même (et donc de la technologie) et caractérisation des modifications de l'aspect de la surface (donc procédé décoratif type engobes etc), ce qui limite la portée de ces travaux. Toutes ces études restent d'ailleurs prudentes sur leurs intentions : il s'agit de démontrer la preuve du concept, et éventuellement d'offrir un outil d'assistance. Il ne s'agit en aucun cas de proposer une méthode de classement totalement automatique. Si on laisse de côté les tentatives précédentes centrées sur la technique de fabrication, il n'existe à première vue que deux études antérieures au projet ARCADIA. Une étude se penchant sur le classement automatique des céramiques au regard des décors [29]. Dans ces travaux, l'Université de Catane s'est intéressée aux fragments de poterie de style de Camarès en Crète. Ce style de céramique, faite au tour 2000 ans avant J.C., est reconnue pour l'extrême finesse de ses parois et la richesse des formes et de la polychromie de ses décorations. Le style de Camarès démontre des combinaisons de couleurs (peinture claire sur fond sombre, blanc sur rouge ou orange, ou sombre sur fond clair) et de motifs recherchant l'harmonie entre la décoration et la forme des récipients : les motifs

curvilignes d'une grande variété (ondulations, spirales, rosettes, rubans, tresses) se répètent et s'entrecroisent, mêlant des thèmes plus figuratifs (palmiers, fleurs, poissons, poulpes, etc). Pour mettre en place un système de classification automatique, l'équipe du Pr. Gallo a collecté une base d'images binaires des éléments de décoration simples figurant dans le corpus des poteries de Camarès. Ces éléments de décoration sont ensuite détectés sur les céramiques puis mis en correspondance avec ceux de la base de référence en se fondant sur une caractérisation de leur contour [30]. L'efficacité de la méthode de détection, pourtant centrale dans leur chaîne, n'est pas vraiment mise en avant dans la publication. Le style décoratif de Camarès est très différent de celui rencontré sur le site de la Médicinerie : nos tessons sont nettement plus rugueux, et les décorations ne sont pas en couleur mais en relief et d'un style bien moins soigné. Le caractère protéiforme des techniques décoratives sur les céramiques, fait que, même si le critère décoratif avait été plus largement traité par analyse d'images, les méthodes qui en découleraient ne seraient pas forcément les plus adaptées aux céramiques de Saran. La deuxième étude menée par Zhou et al. [31] s'intéresse à l'identification de décors gravés sur des tessons avec des tampons. Le décor sur un tesson ne représente qu'une partie du décor du tampon. Leurs travaux visent à assigner un tesson à un tampon. Pour cela chaque tesson est comparé au décor des tampons par des méthodes de curve matching. Nous ne pouvons pas utiliser le même procédé sur nos tessons car nous ne disposons pas de la molette ayant servie à la gravure.

En généralisant la problématique du classement des décors à la molette à celui d'objets pré-industriels fabriqués en série à l'aide d'une matrice, nos tessons peuvent être très raisonnablement rapprochés d'autres types de mobilier archéologique, comme les anciens tampons en bois [32]. Dans ces travaux, l'image de profondeur, créée à partir de la numérisation 3D d'un tampon, est traitée pour obtenir une image binaire aussi proche que possible de l'empreinte laissée par le tampon encré sur une feuille de papier. Pour cette binarisation, l'algorithme de Niblack [33] a été modifié avec un seuil adaptatif. Contrairement à nos tessons, le tampon est parfaitement plat et sans courbure. Le décor en relief est présent sur l'ensemble du tampon, et l'étape de localisation du motif n'est donc pas nécessaire dans ce cas.

Le classement automatique de monnaies anciennes à partir d'images a également été traité dans la littérature [34, 35, 36]. Comme dans notre cas, les auteurs de ces publications ont travaillé sur des chaînes de traitements complètes, depuis l'acquisition jusqu'à la reconnaissance, en faisant face à des pièces de monnaie qui, contrairement à leur contrepartie moderne, présentent une forte variabilité intra-classe. Plusieurs approches on été considérées : utilisation de descripteurs SIFT combinés à un algorithme de flow optique [36, 37], sac-de-mots visuels couplé à un partitionnement spatial [35], mise en correspondance de descripteurs locaux avec vérification de la consistance géométrique des appariements [38], ou encore recalage fin des images dans le domaine spectral et mesure de similarité par corrélation dense [39]. Les développements récents de ces travaux se sont concentrés sur la résolution de problèmes inhérents aux propriétés physiques de ces objets métalliques qui introduisent des artefacts lumineux sur les images, soit en essayant de minimiser leur impact durant la phase d'analyse de l'image [36], soit en proposant une méthode d'acquisition adaptée [34]. Ces développements récents n'apportent donc rien du point de vue des céramiques à la molette. D'autre part, ces travaux de reconnaissance portent sur des pièces de monnaies anciennes qui, même usées, sont généralement entières. Les approches développées ne sont donc pas directement transposables aux objets fragmentaires tels que les tessons de Saran (bien que Marchand prévoyait, à l'origine, d'étendre son système à des carreaux de pavements).

1.3 Contributions

Le projet ARCADIA vise à développer une méthode automatique d'analyse des décors à la molette sur les tessons céramiques trouvés sur le site de la Médecinerie pour faciliter l'interprétation de ce patrimoine archéologique. Cette automatisation doit remplacer la procédure manuelle précédemment effectuée par l'archéologue et devenue trop fastidieuse avec l'augmentation du corpus. L'objectif in fine est de réussir à associer automatiquement les décors à leur molette (ou matrice). Dans ce contexte, la première étape, réalisée au cours de ces travaux de thèse, consiste

à développer une approche complète depuis la numérisation jusqu'à une classification automatique des décors selon leur style de motifs (carré, losange, chevrons, oves, etc). En s'appuyant sur les compétences en céramiques archéologiques de la FAL, les travaux de recherche présentés dans ce manuscrit proposent plusieurs contributions mettant en œuvre des méthodes d'analyse d'images et d'apprentissage automatique.

- Constitution d'une base d'empreintes numériques de tessons :

Une chaine de numérisation des tessons a été mise en place avec un scanner 3D commercial (Next Engine). Environ un millier de tessons ont pu faire l'objet d'un enregistrement (les acquisitions ayant dû être stoppées au cours du projet ARCADIA suite à la fermeture des archives pour travaux de rénovation). Les décors les plus représentés ont été sélectionnés avec l'aide de l'archéologue, en conservant un certain équilibre entre les classes avec une assez bonne représentation de la variabilité des motifs. Les données brutes issues du capteur (nuages de points 3D) ont fait l'objet de pré-traitements pour atténuer le bruit d'acquisition (particulièrement présent à cause de la rugosité de la terre) et palier aux aléas de positionnement du tesson lors de l'acquisition pour générer une base de cartes de profondeur exploitables (images 2D en niveaux de gris).

- Détection automatique des décors :

Cette étape est primordiale pour la réussite de la classification, les zones de décor n'occupant le plus souvent qu'une faible partie du tesson. Elle est rendue complexe par les multiples courbures très variables des tessons, souvent impossible à paramétrer sur des objets aussi fragmentaires. Une méthode originale de détection automatique de la région saillante focalisée sur le décor en relief est proposée.

- Caractérisation des décors :

Contrairement aux approches classiques en indexation mettant en œuvre des descripteurs locaux denses extraits sur des grilles régulières dans les images, un

nouveau descripteur, appelé Blob-SIFT, est proposé pour collecter les signatures (descripteur SIFT) seulement dans les zones pertinentes du motif. Cette approche de grille irrégulière, adaptée à chaque motif, permet à la fois de réduire considérablement la masse de données et d'améliorer les performances de classification. L'efficacité de ce descripteur est avérée par une comparaison à 8 méthodes récentes parmi les plus performantes de l'état de l'art en mettant en œuvre un classifieur SVM ou XGBoost (arbre de décision).

- Approche hybride d'apprentissage par extraction de descripteurs et d'apprentissage profond :

A l'heure où les méthodes d'apprentissage profond prennent le pas sur les approches plus classiques d'extraction de descripteurs associées à un classifieur (SVM ou arbre de décision), nous présentons dans ces travaux une comparaison de la mise en œuvre de ces deux approches. Puis, nous proposons une approche hybride combinant les vecteurs de caractéristiques locales extraites par le descripteur Blob-SIFT et la caractérisation globale du décor fournie par l'apprentissage profond qui améliore encore les performances de classification.

L'organisation du manuscrit est la suivante :

— Le chapitre 2 expose la chaine de numérisation mise en œuvre pour générer les bases de données à partir du matériel archéologique à notre disposition.

— Le chapitre 3 détaille les pré-traitements pour créer les cartes de profondeur à partir des numérisation 3D, puis la méthode de détection automatique de la région saillante du décor.

— Le chapitre 4 présente la classification supervisée par extraction de descripteurs.

— Le chapitre 5 propose une classification mettant en œuvre des techniques d'apprentissage profond.

— Enfin une conclusion termine le manuscrit en ouvrant quelques perspectives à ces travaux de recherche.

Chapitre 2

Matériel archéologique et numérisation 3D

Nous allons expliquer comment les tessons sont référencés ainsi que la méthode que l'archéologue emploie pour extraire les décors des tessons. Et nous détaillerons les compositions des deux bases de données utilisées pour ces travaux.

2.1 Matériel Archéologique

Les fouilles menées par le Service Archéologique Municipal d'Orléans (SAMO) et la Fédération Archéologique du Loiret (FAL) sur le site de La Médicinerie ont mis au jour de nombreux artefacts céramiques dans plusieurs fours (voir Fig. 2.1). Ces morceaux de poteries (tessons) ont fait l'objet d'inventaire et d'analyse par l'archéologue Sébastien Jesset [14]. La taille des tessons est très variée (voir Fig. 2.2). Dans cette étude, nous nous intéressons uniquement aux tessons ayant un décor gravé à la molette. Ce décor est composé d'une répétition de motifs géométriques simples ou composés formant une frise. La répétition d'un motif peut être sur plusieurs lignes, appelées registres (voir Fig. 2.3). Quelle que soit la taille du tesson, le décor est gravé sur 1 millimètre de profondeur environ. Il a une largeur de 1,5 à 3 cm selon les dimensions des cylindres utilisés par les potiers. Chaque tesson est archivé par l'archéologue avec plusieurs informations dont un numéro d'enregistrement, le type de décor observé, la position du décor sur le vase (panse, bord, etc), le cas échéant, la référence à une molette si elle a été identifiée.

Pour une partie des artefacts trouvés, l'archéologue a réalisé des empreintes numériques. La méthode utilisée pour enregistrer les décors présents sur les tessons

FIGURE 2.1 – Artefacts céramiques retrouvés dans un four sur le site de La Médecinerie (Saran, Loiret).

FIGURE 2.2 – Mise en évidence de la variation de la taille des tessons.

comprend 4 étapes : le moulage sur l'argile (Fig. 2.4a), l'encrage (Figure 2.4b), l'impression sur une feuille (Fig. 2.4c), puis la numérisation et la vectorisation pour obtenir une empreinte numérique binaire (Fig. 2.4d). Ce processus manuel prend environ 6 minutes par tesson.

La base de données collectée selon cette procédure contient 1140 empreintes binaires de décors (voir Fig. 2.5). L'archéologue divise cette base en deux caté-

FIGURE 2.3 – Exemple de tesson avec un décor sur deux registres.

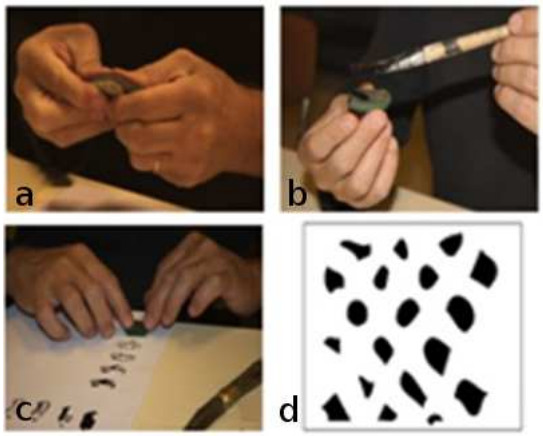


FIGURE 2.4 – Étapes manuelles des empreintes numériques réalisées par l'archéologue : (a) moulage sur l'argile, (b) encrage, (c) impression sur une feuille, (d) numérisation par scanner à plat bureautique.

gories : les décors dits simples (répétition d'un unique motif géométrique) et les décors dits composés (répétition de deux motifs géométriques différents). Chaque type de décor est désigné par une lettre pour les décors simples et par deux lettres pour les décors composés (lettres associées aux deux décors simples). Les tables 2.1 et 2.2 listent les différents décors découverts à Saran. On peut remarquer que les décors les plus courants sont de type losanges, bâtons, carrés sur deux ou trois registres ou chevrons (A,C,G,H,L). Les formes arrondies (oves, classe B) étant plus difficiles à graver sur la molette pour le potier, ce décor est plus rare. Les décors composés sont également très faiblement représentés (moins de 5%).

Il s'agit là d'un premier niveau de classement. Dans une seconde étape, l'archéologue essaye de retrouver les tessons qui ont été réalisés avec la même molette.

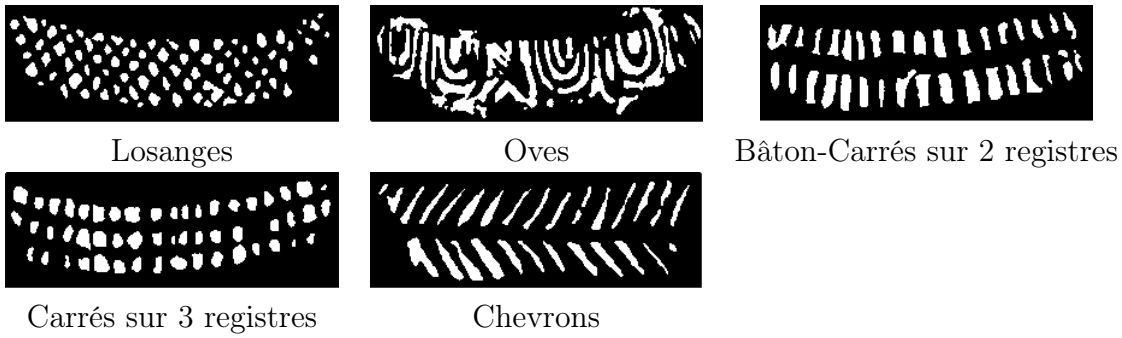


FIGURE 2.5 – Empreintes binaires obtenues par encrage manuel fait par l'archéologue.

TABLE 2.1 – Décors simples

Forme géométrique	**Lettre associée**	**Nombre répertorié**
Losanges	A	150
Oves	B	45
Bâtons	C	61
Triangles	D	8
V	E	6
Carrés sur 1 registre	F	1
Carrés sur 2 registres	G	250
Carrés sur 3 registres	H	453
Croix	I	2
Ondes	J	1
Carrés Ondés	K	1
Chevrons	L	65
Bâtons Entrecroisés	M	3
Médaillons	N	1
Indéterminé	IND	39

Comme il s'agit d'une production artisanale, autant au niveau de la fabrication de la molette, que du point de vue de son application sur la céramique, bien souvent une caractéristique remarquable va permettre à l'archéologue d'identifier la pro-

TABLE 2.2 – Décors composés

Formes géométriques	**Lettres associées**	**Nombre répertorié**
Losanges Triangles	AD	1
Oves Croix	BI	2
Bâtons Triangles	CD	14
Bâtons V	CE	2
Bâtons Croix	CI	14
Triangles V	DE	1
Carrés Chevron	FL	6
Carrés Triangles	GD	4
Carrés Croix	GI	3
Carrés Croix	HI	6
Croix V	IE	1

duction réalisée avec le même outil. Par exemple, un espacement ou une cassure rompant la régularité des motifs qui composent la frise. L'archéologue sélectionne le décor de référence contenant, si possible, l'intégralité du déroulé de la molette. Cependant, l'association des décors à une même molette est un exercice délicat. En effet, l'empreinte numérique obtenue par encrage manuel peut présenter de nombreuses déformations. En premier lieu, lors du marquage du décor sur l'argile fraîche, les différences de pression et d'inclinaison de la molette appliquées par le potier ou les différences de courbure selon les zones d'applications induisent des déformations géométriques. D'autre part, l'outil subi une usure rapide au fil des utilisations. Enfin, le tesson a pu se dégrader lors de son utilisation ou de sa conservation dans le sol.

En second lieu, malgré le soin apporté par l'archéologue lors de l'encrage manuel, les empreintes binaires peuvent présenter des lacunes ou des résidus. Toutes ces déformations entrainent une forte variabilité intra-type et rendent très difficile la classification.

2.2 Numérisation 3D des tessons

Les tessons répertoriés manuellement par l'archéologue ont été numérisés à l'aide du scanner Next Engine [40] acquis par la FAL. Ce scanner a été largement utilisé ces dernières années pour la documentation de patrimoines culturels [41, 42]. Son principal avantage est son faible coût.
L'appareil exploite le principe du profilomètre laser avec un balayage par quatre bandes simultanées pour accélérer le processus de numérisation (voir Fig. 2.6).
La source laser, d'une longueur d'onde de 650 nm (laser rouge), est répertoriée dans la classe 1M, sans danger pour les yeux sauf en cas de vision directe du faisceau.

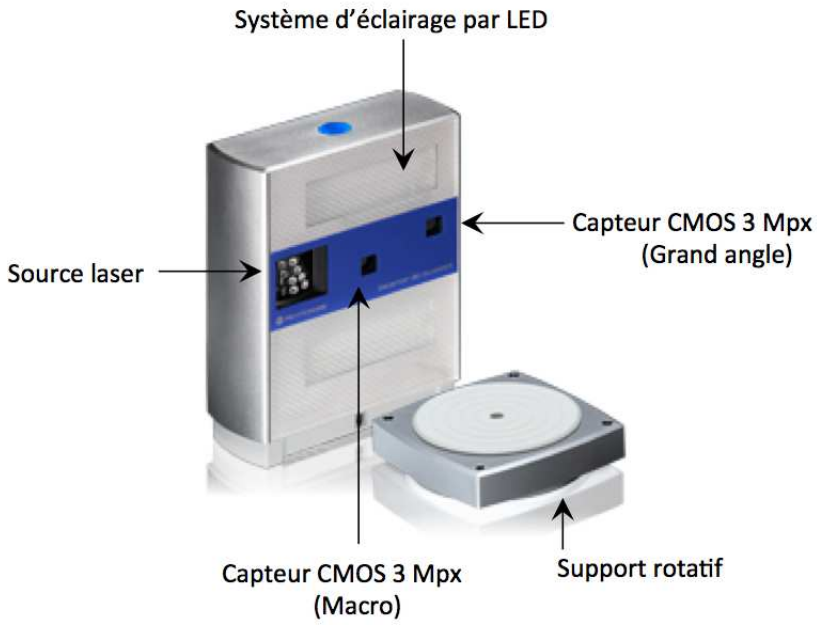


FIGURE 2.6 – Le scanner NextEngine et son plateau rotatif

Le Next Engine est équipé de deux caméras CMOS de résolution 3 Mpx (2048 x 1536 pixels). Le capteur à champ proche (macro) couvre une surface maximale de 12,9 x 9,6 cm pour une distance de travail optimale de 16,5 cm. Ce capteur permet de travailler sur des petits objets comme nos tessons. Le capteur grand angle permet de numériser des objets de taille plus importante, de 34,3 x 25,6 cm, avec une distance de travail optimale de 43,2 cm.

La source laser, alliée à l'un ou l'autre des capteurs, permet, selon le constructeur, d'obtenir deux niveaux de densité de points : 25000 points par cm2 (ppcm2) pour le capteur macro, contre 3500 ppcm2 pour le grand champ. Le constructeur annonce une précision de 0,0127 cm pour le capteur macro. Mais l'information de précision donnée par le constructeur semble être liée à la résolution angulaire (le plus petit écart entre deux points). Une publication évalue d'ailleurs la reproductibilité du montage à ±0,084 cm [43], soit bien au-dessus des chiffres avancés par NextEngine.

Les caméras permettent d'acquérir la texture trichromatique de l'objet (RGB) ou multispectrale (7 longueurs d'ondes non spécifiées par le constructeur). Le scanner dispose également d'un système d'éclairage par LED pour l'illumination de l'objet, sans avoir à recourir à une source lumineuse extérieure.

Dans le cadre de notre étude, nous nous intéressons uniquement au décor. Ainsi, seule la face présentant le décor est numérisée. Un tesson est donc scanné par un unique balayage (voir Fig. 2.7). Le temps d'acquisition est ainsi réduit à une minute et demie environ par tesson.

Le scanner délivre un nuage de points 3D enregistré dans un fichier informatique (nous avons retenu le format OBJ). Ce fichier représente les coordonnées X,Y,Z des points 3D dans un repère associé à la caméra sélectionnée. L'axe des Z est l'axe optique orienté vers le scanner. Les axes X, Y sont parallèles au plan image et évoluant respectivement vers la droite et vers le haut. La documentation technique du scanner ne précise pas l'origine du repère, dont on peut supposer qu'elle se situe au centre optique de la caméra. Le scanner délivre aussi un fichier image au format JPEG et un fichier au format MTL (Material Template Library) permettant de relier l'information 3D et l'information colorimétrique de l'image JPEG.

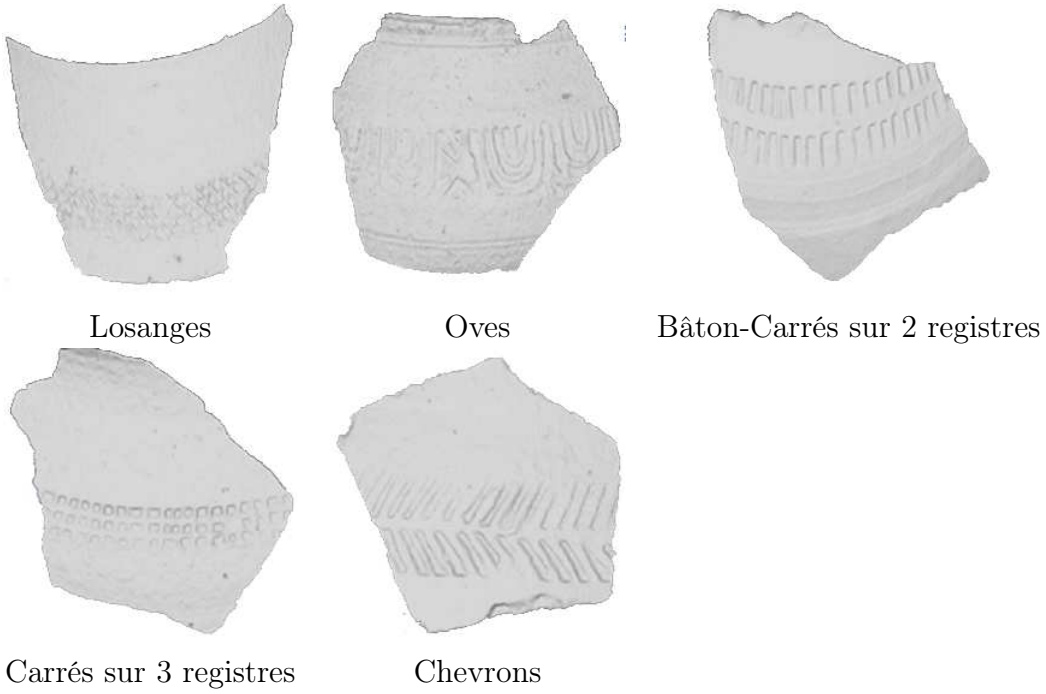


FIGURE 2.7 – Exemples de scans de tessons de classes différentes.

2.3 Bases de données

Parmi les 1140 tessons enregistrés par l'archéologue, nous avons conservé les décors uniquement suffisamment représentés. Les classes D, E, F, I, J, K, M et N ont été écartées, tout comme les décors composés. La classe indéterminée n'est pas prise en compte car le type de décor n'est pas connu. Une étude de faisabilité a exploité une première base [44, 45], constituée de 377 empreintes binaires et les scans 3D associés (voir table 2.3). Les bâtons et les carrées sur deux registres sont très ressemblant comme le montre les figures 2.8 et 2.9 et ont été fusionnés en une classe unique (C-G).

FIGURE 2.8 – Décor classé dans la classe carrés sur deux registres.

Une nouvelle campagne de numérisation a permis d'augmenter le corpus pour

FIGURE 2.9 – Décor classé dans la classe bâtons.

TABLE 2.3 – Base 1 utilisée pour l'étude de faisabilité (377 tessons)

Classes	A	B	C-G	H	L
Nombre de tessons	64	24	117	144	28

les quatre classes les plus représentées (A,H,L,C-G). Les décors de type B (oves) restent très minoritaires sur les tessons découverts à Saran et ont été écartés pour conserver un certain équilibre entre les classes de la base (voir table 2.4). La base 2 inclut donc toutes les données de la base 1 à l'exception de la classe B (oves). Pour cette base nous ne disposons pas des empreintes manuelles réalisées par l'archéologue. Nous disposons des cartes de profondeurs (voir Fig. 2.10), des cartes de variances locales (voir Fig. 2.11) et des décors binaires (voir Fig. 2.12). Dans le chapitre suivant, les méthodes pour obtenir ces images seront expliquées.

TABLE 2.4 – Base 2 utilisée pour tester nos algorithmes (888 tessons)

Classes	A	C-G	H	L
Nombre de tessons	211	259	274	144

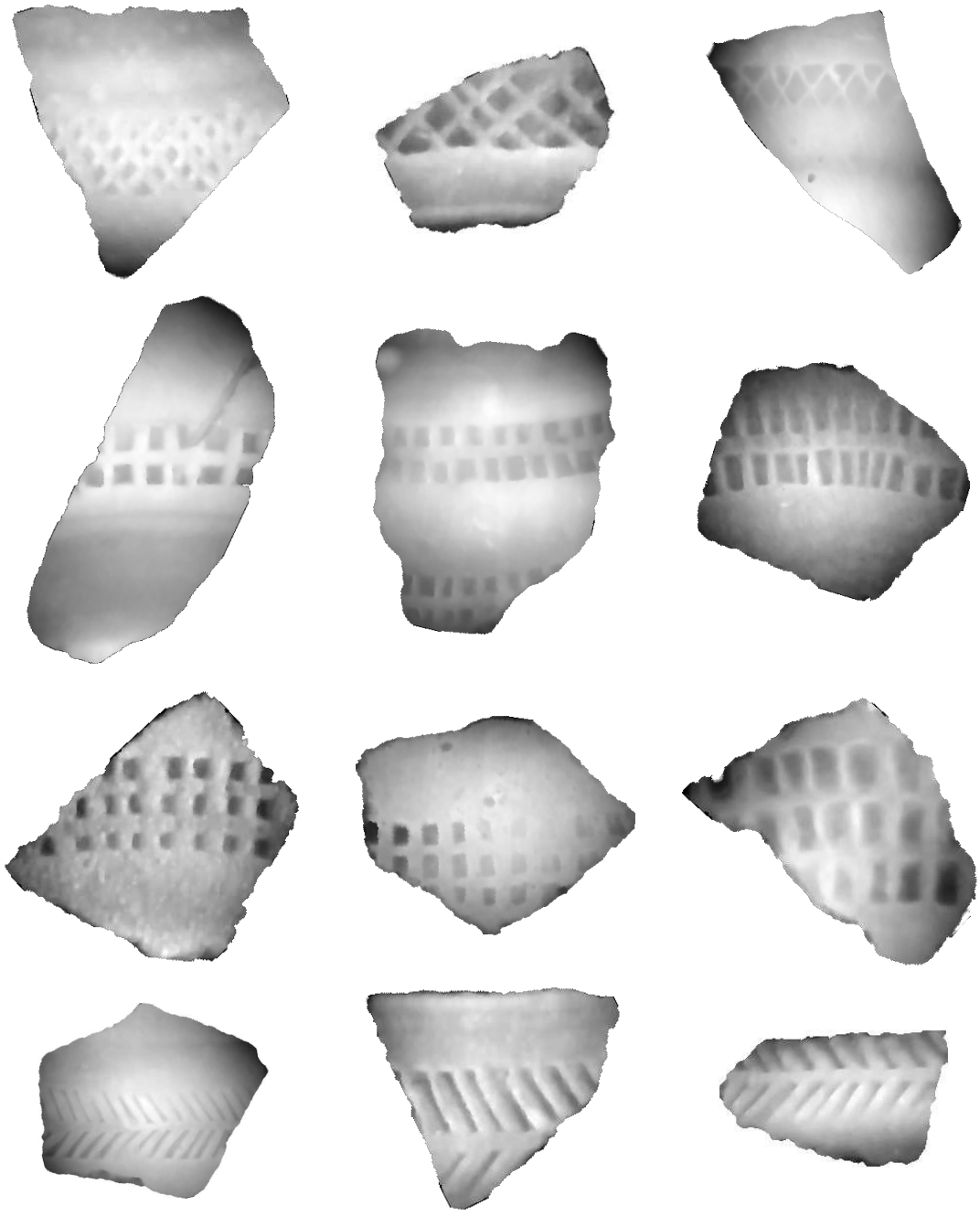


FIGURE 2.10 – Exemples de cartes de profondeurs.

FIGURE 2.11 – Exemples de cartes des variances locales.

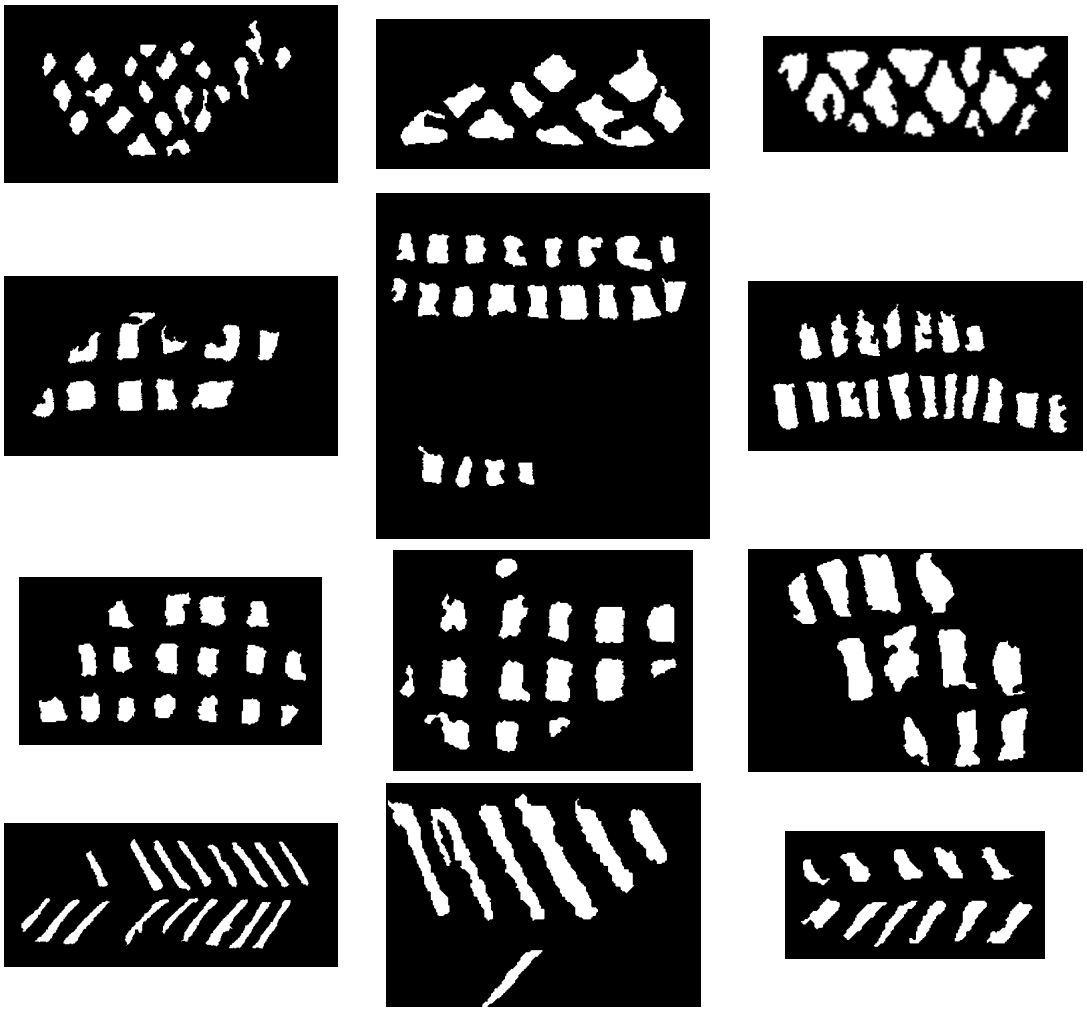


FIGURE 2.12 – Exemples de décors binarisés.

Chapitre 3

Prétraitements et extraction du décor

Comme nous cherchons à classer les tessons par type de décor, nous allons extraire la ou les parties du tesson contenant le ou les décor(s) pour éviter que les artefacts influent sur la classification.

Les données brutes issues du capteur font l'objet de prétraitements afin d'atténuer notamment le bruit d'acquisition et palier aux éventuels aléas de positionnement du tesson devant le capteur ou de courbure du tesson. Le nuage de points est ainsi transformé en carte de profondeur 2D et une méthode efficace a été développée pour détecter automatiquement les régions saillantes focalisées sur le décor en relief. La figure 3.1 résume les étapes effectuées. Les sections suivantes détaillent les traitements.

3.1 Carte de profondeurs

Le scanner délivre un fichier contenant la liste des coordonnées des points 3D qui sont généralement bruitées. On applique un filtrage permettant de supprimer les points aberrants (outlier removal) basé sur la distribution locale des points (PCL [46]). Ce filtre considère la moyenne et l'écart-type de la distance de chaque point à ses voisins. En supposant que la distribution résultante est gaussienne, tout point qui s'écarte de son voisinage de plus d'un écart-type de la distance moyenne est éliminé.

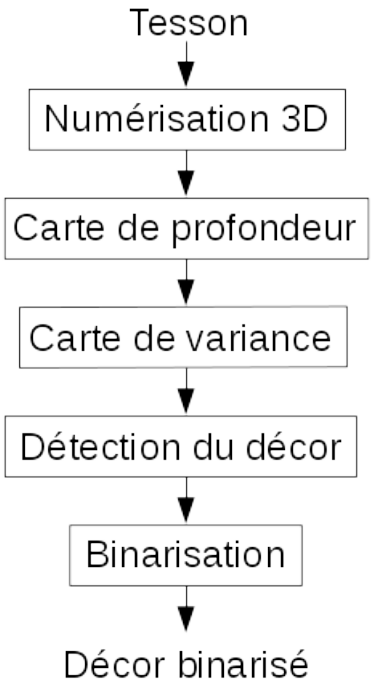


FIGURE 3.1 – Chaine de traitements proposée pour la binarisation des décors.

L'organisation en nuage de points 3D ne facilite pas les traitements de voisinage. Il est possible de transformer le nuage en une image de 256 niveaux de gris appelée carte de profondeur où chaque pixel représente le relief. L'intérêt est de pouvoir appliquer les traitements d'images conventionnels (filtrage, segmentation, etc). Comme le relief des tessons est faible comparé à la distance moyenne au scanner, nous avons effectué une projection orthographique pour construire la carte de profondeur. Au préalable, on applique une transformation rigide pour compenser les problèmes de centrage et d'orientation des tessons lors de la numérisation 3D. Une Analyse en Composantes Principales (ACP) sur les coordonnées (X, Y, Z) délivrées par le scanner permet de récupérer un repère centré sur le tesson par le centroïde du nuage de points 3D et ses directions principales. Cette ACP permet non seulement de redresser le tesson visuellement, mais aussi de définir le plan de projection sur la base des deux premières dimensions principales. L'épaisseur du tesson étant plus faible par rapport à la surface, et la troisième dimension principale étant orthogonale aux deux premières, la troisième dimension sera orientée suivant la direction de l'épaisseur du tesson. Les deux premières dimensions formeront un plan de projection équilibré par rapport à la courbure du tesson. La transformation rigide (rotations et translations) effectue le changement de repère

pour passer des coordonnées du scanner aux coordonnées centrées sur le tesson (X_t, Y_t, Z_t). Une projection directe est ensuite effectuée : le niveau de gris assigné à chaque pixel (X_t, Y_t) est proportionnel à la profondeur Z_t, pour créer la carte de profondeur.

Du fait du balayage unique par le scanner lors de la numérisation de la face du tesson, il arrive qu'il y ait des données lacunaires dans le nuage de points 3D. Ainsi, certaines «taches noires» sont visibles dans les zones d'occultation des faisceaux laser projetés (voir la partie gauche de la Fig. 3.2a). Pour combler ces lacunes, ces «taches noires» sont automatiquement détectées dans la carte de profondeur et un algorithme d'inpainting [47] estime l'information manquante. Cet algorithme utilise un estimateur de lissage (moyenne pondérée du voisinage) se propageant le long du gradient de l'image. Une méthode de level-set (fast marching [48]) est appliquée pour détecter les contours des régions manquantes et compléter l'image. Enfin, une égalisation de l'histogramme permet d'augmenter le contraste de la carte de profondeur. Ce qui permet de mieux répartir les intensités sur l'ensemble de la plage de valeurs possibles, dans notre cas de 0 à 255. L'effet résultant est montré sur la figure 3.2b. Afin de mettre en évidence le relief peu profond du décor, un estimateur de variance locale est appliqué à cette carte de profondeur (voir Fig. 3.2c). Cet estimateur est détaillé dans la section suivante.

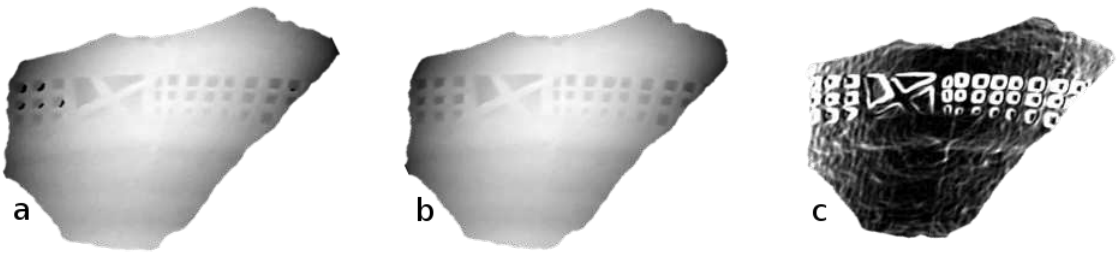


FIGURE 3.2 – (a) Projection du nuage de points 3D, (b) amélioration de la qualité de la carte de profondeur par l'application de la méthode d'inpainting, (c) carte de variance locale.

3.2 Carte de variance locale

L'opérateur de variance locale agit comme un filtre passe-haut pour atténuer la courbure du tesson. Comme le montre la figure 3.3, la variation du niveau de gris dans la carte de profondeur est due principalement à la courbure du tesson car la variation de profondeur des décors est faible (environ 1 millimètre). Une première approche consisterait à estimer la surface courbe du tesson pour la soustraire. Cependant, il est difficile d'estimer avec précision les courbures de fragments de si petite taille car le bruit d'acquisition causé par la surface granuleuse de l'argile est élevé par rapport au relief du décor. L'application du détecteur de Canny [49], par exemple, n'est pas efficace dans ce cas car les contours du décor ne sont pas différentiables de la rugosité de l'argile. La variance locale est calculée comme suit :

$$V_{x,y}^2 = \frac{\sum_{i=0}^{n}(Ii_{x,y} - \mu)^2}{n} \tag{3.1}$$

avec : $V_{x,y}^2$: la valeur de la variance du pixel de coordonnées (x, y).
$Ii_{x,y}$: la valeur du niveau de gris du pixel de coordonnées (x, y).
n : le nombre de pixel dans la fenêtre, ici 3x3 pixels.
μ : la valeur de niveau de gris moyen de la fenêtre.

Comme on peut le voir sur la figure 3.4, les contours du décor sont bien plus visibles sur la carte des variances locales que sur la carte de profondeur (Fig. 3.3).

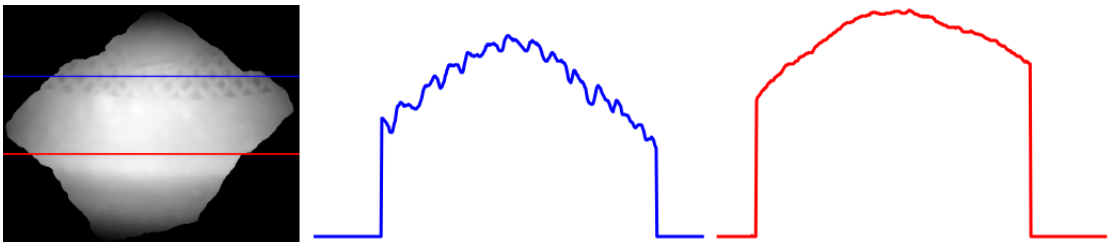


FIGURE 3.3 – Profil des niveaux de gris sur deux lignes de la carte de profondeur, une contenant un décor (bleu) et une sans (rouge).

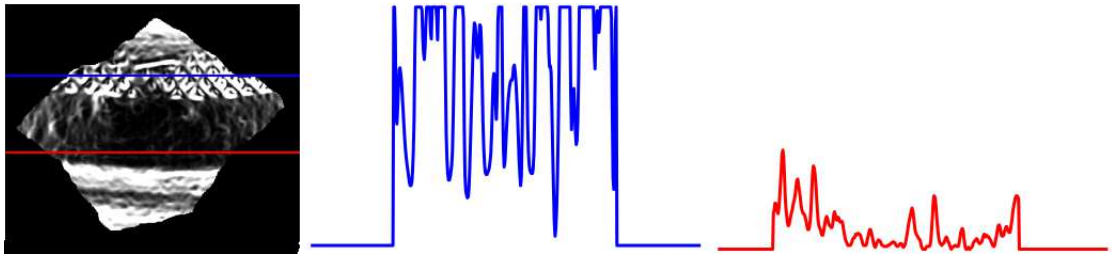


FIGURE 3.4 – Profil des niveaux de gris sur deux lignes de la carte des variances locales, une contenant un décor (bleu) et une sans (rouge).

3.3 Détection des régions saillantes

Une binarisation des cartes de profondeurs de tout le tesson par un seuillage adaptatif ne permet pas d'isoler le décor. L'idée est alors d'isoler les régions saillantes qui correspondent aux motifs géométriques, donc aux zones les plus texturées. La méthode proposée consiste à appliquer un détecteur de points singuliers qui s'accrochent aux motifs géométriques, puis à rechercher les zones de forte concentration de points sur la carte des variances locales. Parmi les détecteurs de points existant dans la litérature (Harris [50], SIFT [2], SURF [51], ORB [52], BRISK [53], MSER [54]), FAST (Features from Accelerated Segment Test) [1] est celui qui offre le meilleur compromis entre le nombre de points et la vitesse de détection sur nos images. Comme dans l' article original [1], nous avons considéré un voisinage circulaire de 16 pixels (représenté par une ligne pointillée blanche sur la figure 3.5). Le pixel central est déclaré comme un point caractéristique si au moins 12 pixels contigus autour du cercle ont une différence d'intensité avec ce point central supérieure (ou inférieure) à un certain seuil. Pour accélérer le traitement, le premier test ne considère que les intensités des quatre pixels cardinaux (1, 5, 9 et 13) et élimine les pixels qui ne sont pas un coin. Ensuite, le critère de test complet est appliqué aux candidats potentiels restants.

La figure 3.6 montre l'influence du seuil de différence appliqué sur le nombre de points FAST détectés. Si le seuil est trop bas, les points caractéristiques seront répartis sur l'ensemble du tesson accrochant la rugosité de l'argile (voir Fig. 3.6a). Au contraire, si le seuil est trop élevé, il n'y aura pas assez de points sur la région de saillance pour un regroupement efficace (voir Fig. 3.6c). La figure 3.6b

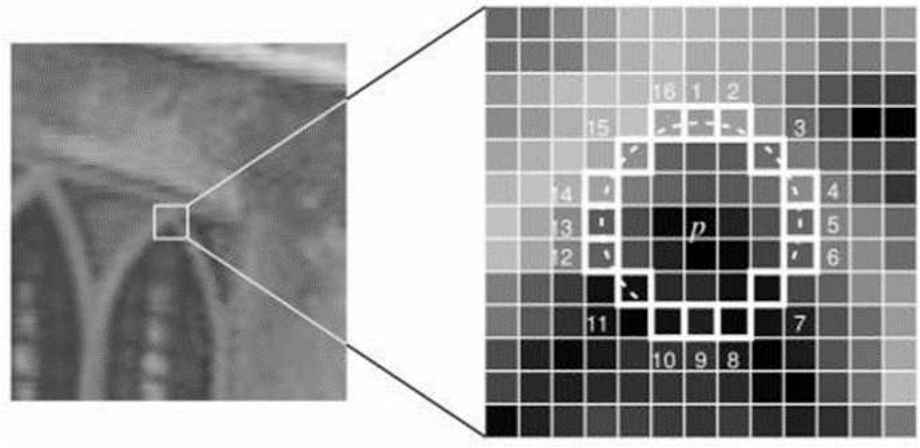


FIGURE 3.5 – Voisinage utilisé pour déterminer si un pixel est un point caractéristique ou non avec le détecteur FAST (source [1])

montre un résultat correct. Nous avons effectué des tests expérimentaux sur notre ensemble de données pour déterminer ce paramètre (voir section 3.5.2).

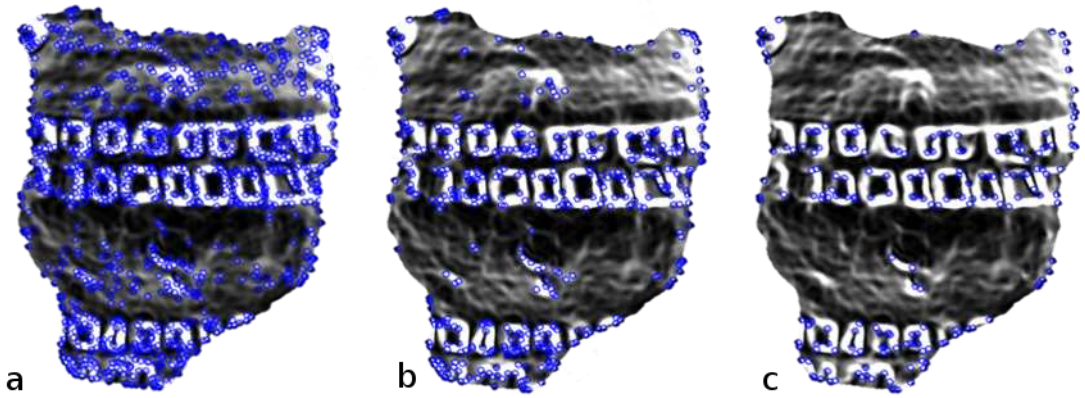


FIGURE 3.6 – Points FAST détectés sur la carte des variances locales avec un seuil de 20 (a), 41 (b), 56(c).

Pour extraire les régions saillantes correspondantes aux zones de forte concentration de points singuliers, on a utilisé la méthode de regroupement de points appelée Density-Based Spatial Clustering of Applications with Noise (DBSCAN) [55]. Un des avantages de cette méthode est qu'il n'est pas nécessaire de fixer le nombre de groupes (clusters), qui peut varier d'un tesson à l'autre. La méthode fonctionne avec deux paramètres : le nombre minimum de points requis pour créer un cluster, nc_{min} et la distance maximale entre deux points appartenant à un même groupe, d_{max}. Dans notre cas, ces deux paramètres sont liés à la densité de points

détectés sur chaque tesson, comme suit :

$$d_{max} = \sqrt{\frac{nc_{min}}{D\pi}} \tag{3.2}$$

avec D : la densité de points globale obtenue par le rapport $\frac{n}{s}$ où n est le nombre de points FAST détectés et s la taille du tesson (en pixels).

La valeur du paramètre nc_{min} sera justifiée à la section 3.5.3. De cette façon, les points isolés sont éliminés et seules les zones de plus grande densité de points FAST sont sélectionnées pour extraire le contour global du décor comme on peut le voir sur la figure 3.7.

FIGURE 3.7 – Détection de zones de points denses par DBSCAN.

Ensuite le contour du cluster est obtenu par triangulation de Delaunay [56]. La forme du cluster peut être une enveloppe non convexe, pour avoir la forme la plus proche possible du contour des décors (voir un exemple sur la figure 3.8).

Comme on peut le voir sur la figure 3.9a, le regroupement par DBSCAN ne permet pas toujours de couvrir l'intégralité du décor car la densité de points peut diminuer subitement dans certaines zones, en raison d'un changement de motif ou de l'érosion des tessons. Afin de surmonter cette difficulté, nous avons exploité l'alignement des décors induit par l'application de la molette pour regrouper et remplir automatiquement les zones manquantes. Deux méthodes ont été testées. La première approche exploite une analyse en composantes principales (ACP) sur les coordonnées des points FAST. La direction de l'application de la molette en

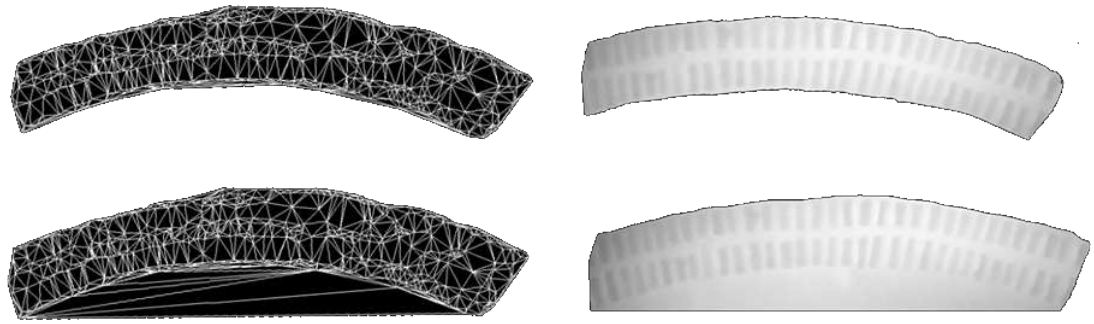


FIGURE 3.8 – Triangulation de Delaunay avec (haut) et sans (bas) déconnection des sommets éloignés et la région de saillance correspondante.

bois peut être détectée comme la première composante, tandis que la deuxième composante est perpendiculaire à celle-ci. Les projections orthogonales des points selon la direction de la première composante auront un fort chevauchement si elles appartiennent à deux clusters alignés. Nous pouvons facilement fusionner ces deux clusters comme appartenant au même décor.

Une seconde approche, Parallel Clusterwise Linear Regression (P-CLR), plus générale a été développée par classification non supervisée [57]. L'idée est d'imposer une forme commune à tous les clusters caractérisée par un hyperplan qui sera le même pour tous les groupes à une translation près. Les points sont donc supposés être distribués autour d'hyperplans parallèles. La fonction objectif utilisée s'exprime comme la minimisation de la somme des distances de chaque point à son hyperplan. Comme pour le cas de K-means [58], la résolution est effectuée par l'alternance de phases d'affectation de chaque point à l'hyperplan le plus proche et de phases de calcul de l'hyperplan qui ajuste au mieux l'ensemble des points qui lui sont affectés. L'objectif étant d'obtenir des hyperplans parallèles, cette phase de calcul est menée simultanément pour tous les hyperplans, par une méthode de régression.

Une comparaison des deux méthodes sera proposée en section 3.5.4. Elles ont une efficacité équivalente, la méthode P-CLR étant plus sensible à l'initialisation. La méthode basée sur une ACP est retenue aussi pour sa facilité d'implémentation. La figure 3.9b montre l'étape de fusion des clusters et la zone de saillance extraite

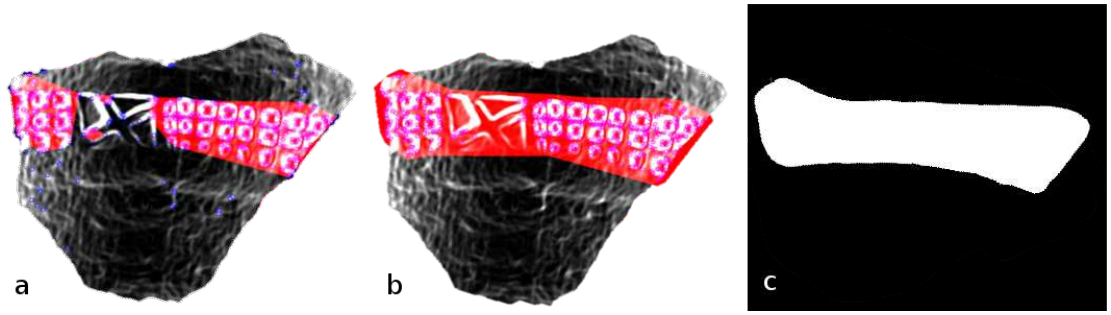


FIGURE 3.9 – (a) Clusters calculés par DBSCAN à partir des points FAST détectés sur la carte des variances locales, (b) fusion des clusters alignés, (c) création du masque représentant la ou les zones saillantes du tesson.

où le décor est localisé(voir Fig. 3.9c). Ainsi la binarisation du décor sera appliquée uniquement dans la zone de saillance.

3.4 Binarisation du décor

Pour obtenir une image binaire semblable aux empreintes réalisées par l'archéologue, l'algorithme de Niblack [33] est appliqué sur la carte de profondeur (uniquement dans la zone de saillance). Un seuil adaptatif est calculé sur une fenêtre glissante de taille 5x5 centrée sur le pixel central par :

$$S(i,j) = \mu_w + k\sigma_w \tag{3.3}$$

avec :

μ_w : le niveau de gris moyen dans la fenêtre.

σ_w : l'écart type moyen des niveaux de gris dans la fenêtre.

k : une constante fixée à 0.2.

Même en variant la valeur de k, cette binarisation révèle des artefacts (voir table 3.1a). Comme la surface des artefacts varie selon les décors, nous avons choisi un critère de taille adaptatif : toutes les particules dont la surface est inférieure à la moitié de la surface moyenne des éléments du décor sont éliminés (voir table 3.1b). Notons que la version améliorée du seuillage Niblack proposée par Sauvola [59] pour réduire le bruit de fond n'est pas efficace dans notre cas car certains éléments

sont pris comme artefacts et sont supprimés.

On constate aussi des connexions indésirables entre les éléments. Elles peuvent être supprimées en utilisant des propriétés géométriques (voir table 3.1c). Une simple érosion morphologique fausserait la forme des motifs composant le décor. Pour éviter cela, chaque motif élémentaire est délimité par une boîte englobante rectangulaire dont la direction principale est calculée en appliquant une ACP. Le nombre de pixels blancs est sommé ligne à ligne. Si la somme des pixels blancs sur une ligne est inférieure au quart de la longueur du motif, le niveau de gris de tous les pixels de cette ligne est mis à zéro. La même opération est appliquée aux colonnes. Les motifs étant principalement des bâtons, des carrés, des chevrons et des losanges, cette technique est relativement efficace pour supprimer les connexions indésirables. Comme on le voit sur la figure 3.1c, les motifs qui étaient connectés ne le sont plus. La table 3.1d montre que le décor obtenu par la chaine d'acquisition automatique est moins bruité que la version d'encrage manuel effectuée par l'archéologue. Le décor est maintenant exploitable pour être caractérisé.

TABLE 3.1 – Étapes de binarisation du décor dans la région de saillance.

	Tesson n° 17	**Tesson n°141**
(a) Binarisation de Niblack		
(b) Suppression des artefacts		
(c) Suppression des connexions indésirables		
(d) Empreinte manuelle		
(e) Visualisation 3D		

3.5 Réglages des paramètres et validation des traitements

Cette section justifie le réglage du paramètre de seuil pour le détecteur FAST, ainsi que celui du minimum de points pour former un cluster et détecter les zones de décor par DBSCAN. Enfin, une troisième partie présente les tests réalisés avec le jeu de données de la base 2 (888 tessons) pour valider cette étape d'extraction des décors au regard d'une vérité terrain.

3.5.1 Métrique d'évaluation

Pour évaluer quantitativement la qualité des régions de saillance extraites automatiquement, nous les avons comparées avec celles extraites manuellement (vérité terrain). La figure 3.10 illustre les métriques de recouvrement utilisées pour cette évaluation. La zone blanche représente la vérité terrain (VT). On considère alors 3 zones différentes : la zone de chevauchement entre la région détectée automatiquement et la VT (zone 1), les pixels non détectés de la VT qui sont de faux négatifs (zone 2) et les pixels mal détectés en dehors de la VT qui sont des faux positifs (zone rouge 3). La VT se compose donc de la zone 1 et de la zone 2, la région de saillance de la zone 1 et de la zone 3. Chaque zone a été évaluée par le nombre de pixels inclus. Ensuite, le pourcentage des zones de bonnes détections a été estimé par $\frac{zone1}{zone1+zone2}$ et le pourcentage des zones de détection manquées par $\frac{zone2+zone3}{zone1+zone2+zone3}$.

3.5.2 Seuil pour le détecteur FAST

Le seul paramètre du détecteur FAST à fixer est le seuil sur la différence des niveaux de gris. Ce paramètre affecte fortement le nombre de points caractéristiques détectés sur la carte des variances locales et donc la région de saillance qui sera extraite après le regroupement des points par DBSCAN.

Comme le nombre de points à détecter peut varier considérablement d'un tesson à l'autre en fonction de sa taille, nous avons adopté un seuil adaptatif. Nous

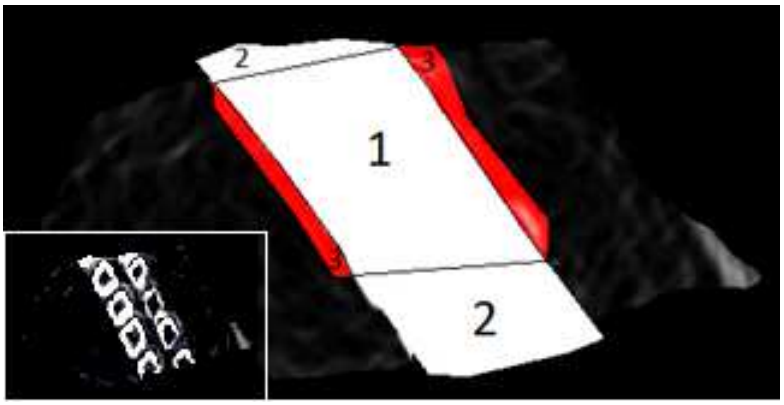


FIGURE 3.10 – L'évaluation du chevauchement entre la zone de saillance détectée automatiquement et la VT manuelle (zone blanche) : (1) est la zone de recouvrement, (2) les pixels non détectés, c'est-à-dire les faux négatifs et (3) la zone rouge contient les pixels détectés incorrectement, c'est-à-dire les faux positifs.

avons appliqué le détecteur FAST en augmentant la valeur de seuil de différence progressivement et mesuré l'effet de ce seuil sur les taux de chevauchement avec la vérité terrain (voir section 3.5.1). Nous en avons déduit qu'un seuil adaptatif conservant 17% du nombre maximum de points détectés sur chaque tesson donne le meilleur résultat (le nombre maximum est donné pour un seuil égal à 1). A titre d'illustration, la figure 3.11 donne l'évolution du nombre de points caractéristiques détectés en fontion du seuil pour le même tesson présenté à la figure 3.6, et la valeur finalement sélectionnée pour ce tesson.

3.5.3 Taille des clusters pour DBSCAN

L'étape de clustering dépend du nombre minimum de points considérés pour former un cluster. La figure 3.12 montre l'effet de ce paramètre sur la région saillante détectée (contour rouge). La meilleure détection peut être observée à la figure 3.12b. Comme précédemment, nous avons testé différentes valeurs pour ce paramètre sur l'ensemble des données (base 2 : 888 images) et nous avons observé qu'en fixant le nombre minimum de points à 11, nous obtenons le meilleur résultat pour la détection des régions saillantes par rapport à la vérité terrain obtenue manuellement (voir section 3.5.1).

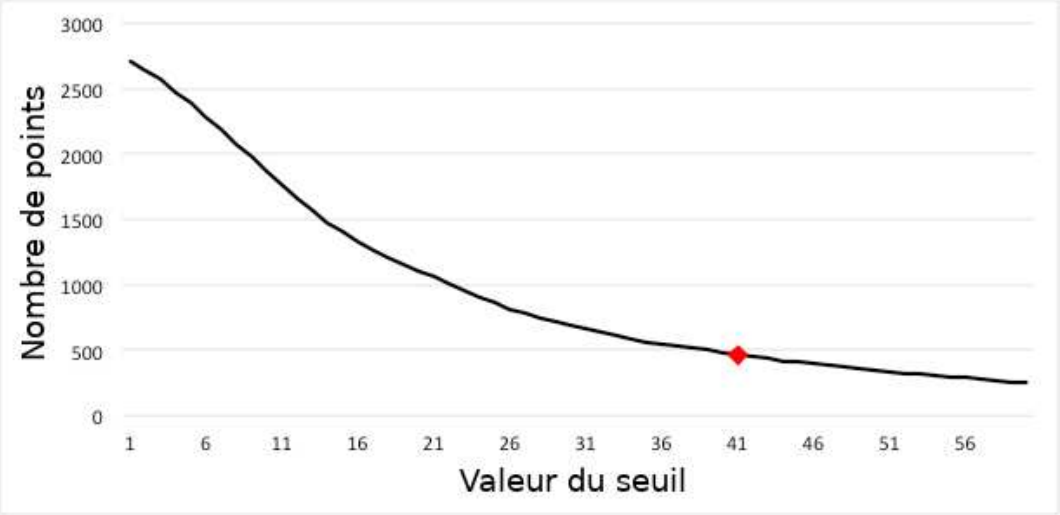


FIGURE 3.11 – Variation du nombre de points caractéristiques détectés en fonction du seuil de différence. Le seuil adaptatif finalement sélectionné pour ce tesson est de 41 correspondant à 462 points caractéristiques détectés par FAST (17% du maximum 2718 points).

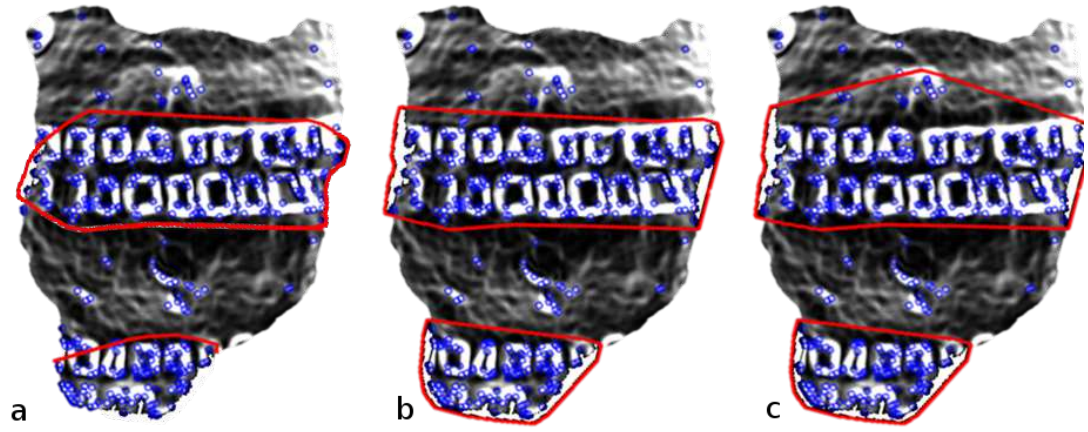


FIGURE 3.12 – Influence du nombre de points minimum pour former un cluster pour détecter les régions saillantes (lignes rouges) : ce nombre est fixé à 3 (a), 5 (b), 7 (c).

3.5.4 Comparaison des méthodes de recherche d'alignement des décors

La table 3.2 présente la comparaison des deux méthodes P-CLR et ACP sur la base 1 (377 tessons).

Tant pour les taux de bonne et de mauvaise détections, l'efficacité des deux méthodes est équivalente. Mais la méthode P-CLR dépend de l'initialisation ; la fonction optimisée pouvant présenter, selon le jeu de données considéré, de nombreux optimums locaux. C'est donc la méthode ACP qui est conservée.

TABLE 3.2 – Détection de l'alignement des décors : comparaison des deux méthodes.

Methode	**Taux de bonne détection**	**Taux de mauvaise détection**
P-CLR	90.40	32.40
ACP	90.02	33.24

3.5.5 Évaluation de l'extraction des régions saillantes

La table 3.3 présente la moyenne obtenue sur la base 2 des métriques définies à la section 3.5.1 : taux de bonne détection et taux de mauvaise détection. Pour ces tests, le seuil adaptatif pour le détecteur FAST est fixé à 17%. Les taux moyens de bonne détection sont supérieurs à 90% sur une large gamme de 7 à 37 points minimum pour la formation d'un cluster. Sur cet intervalle, le taux de mauvaise détection oscille autour de 41-42%. On remarque que le nombre de points minimum pour former un cluster n'est pas un paramètre très sensible. Le nombre de 11 est un bon compromis qui permet d'obtenir un taux de bonne détection de 91.18% et un taux de mauvaise détection de 41.08%. Sur les mauvaises détections, on a pu observer que les parties du décor non détectées (zone 2) représentent 8.81% et les faux positifs (zone 3) représentent 32.27%. Pour toutes les expérimentations ultérieures, le seuil adaptatif du détecteur FAST sera fixé à 17% et le nombre de points minimum pour former un cluster nc_{min} à 11.

TABLE 3.3 – Les taux de bonne et mauvaise détections par rapport à la vérité terrain manuelle ont été calculés sur la base de 888 tessons en faisant varier le nombre de points minimum pour former un cluster.

Nombre minimum de points de points pour former un cluster	**Taux moyen de bonne détection (%)**	**Taux moyen de mauvaise détection (%)**
3	84.25	40.94
5	88.95	40.88
7	90.41	40.86
9	90.83	41.13
11	**91.18**	**41.08**
13	91.16	41.34
15	91.14	41.34
17	90.97	41.48
27	91.00	42.13
37	90.98	42.90
47	89.89	43.84
57	88.78	45.13
67	87.91	45.99
97	83.94	49.93

Chapitre 4

Apprentissage par extraction de descripteurs

Les zones saillantes étant isolées, nous allons maintenant caractériser les décors des tessons par l'utilisation de descripteurs. L'extraction de descripteurs permet de passer d'une image à un vecteur de caractéristiques pour identifier le décor. Un descripteur doit être à la fois discriminant entre les classes et robuste à la variabilité intra-classe. Si les objets à identifier ne sont pas redressés ou recalés, les descripteurs doivent également être invariants aux transformations d'images. Dans notre cas, les descripteurs peuvent être extraits à partir des cartes de profondeurs, des variances locales ou sur les images binaires.

Ce chapitre présente les descripteurs et classifieurs retenus, puis les tests réalisés sur nos bases de décors.

4.1 Descripteurs

Les décors de nos tessons sont des frises composées de motifs répétitifs. Les descripteurs retenus doivent caractériser le motif élémentaire et sa répétition.
On distingue deux types de descripteurs dans la littérature : les descripteurs globaux calculés sur toute l'image et les descripteurs locaux calculés autour de points d'intérêt.

Une étude antérieure [60] a évalué plusieurs descripteurs globaux sur la base 1 (377 images) : les matrices de cooccurrence (GLCM), les filtres de Gabor [26],

les covariogrammes, les LBP [61] et les moments de Zernicke [62]. L'outil d'analyse Explorer3D [63], a révélé que les filtres de Gabor étaient les plus efficaces dans notre contexte. Ce logiciel permet d'opérer et de visualiser une réduction de dimension, donc de visualiser la distribution spatiale des objets en fonction de différents jeux de descripteurs. Une bonne distribution spatiale (classes cohérentes et bien séparées) est alors un indice de pertinence des descripteurs. Comme le montre la figure 4.1, les descripteurs de Gabor offrent une bien meilleure séparation des classes dans l'espace. La représentation spatio-fréquentielle des filtres de Gabor se prête bien à la caractérisation de nos décors répétitifs et sera celle finalement retenue comme descripteur global.

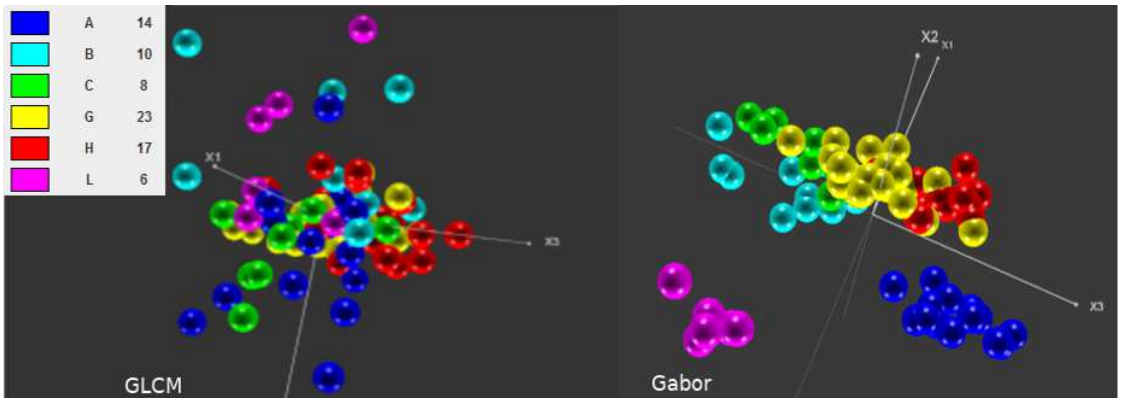


FIGURE 4.1 – Visualisation avec Explorer3D des descripteurs GLCM (à gauche) et Gabor (à droite) sur une base de 78 tessons après projection sur les 3 premières composantes principales.

Les descripteurs locaux autour de points d'intérêt ont rencontré un vif succès ces dernières années pour l'indexation d'images ou d'objets. Leur force réside dans leur robustesse à l'occultation partielle et dans une certaine mesure aux transformations géométriques. L'extraction d'un grand nombre de descripteurs locaux permet d'élaborer un dictionnaire visuel.

L'approche dite par sac-de-mots est une méthode d'apprentissage supervisé qui se montre très performante dans la littérature. Nous proposons donc de la tester sur nos décors avec différents types de descripteurs : le détecteur SURF [51], le descripteur robuste SIFT [2] sur une grille régulière ou au travers d'une pyramide spatiale PHOW [64] et enfin une adaptation à une grille irrégulière induite par les

motifs.

Les sections suivantes détaillent les descripteurs retenus.

4.1.1 Filtres de Gabor

Un filtre de Gabor est une onde plane sinusoïdale modulée par une enveloppe gaussienne :

$$g(x,y) = exp(2j\pi(u_0x + v_0y) + \phi) \cdot exp(-(\frac{(x-x_0)^2}{\sigma_x^2} + \frac{(y-y_0)^2}{\sigma_y^2})) \tag{4.1}$$

Il peut être pratique de la voir comme un couple de fonctions réelles, déphasées de $\frac{\pi}{2}$. Il s'agit alors de la partie réelle et la partie imaginaire de la fonction complexe :

$$G1(x,y) = \cos(ax + by) \cdot exp(-\frac{x^2+y^2}{2\sigma^2}) \tag{4.2}$$

$$G2(x,y) = \sin(ax + by) \cdot exp(-\frac{x^2+y^2}{2\sigma^2}) \tag{4.3}$$

où les variables a et b déterminent la fréquence et l'orientation du filtre, le terme σ^2 son étendue en modifiant la variance de la gaussienne.

Les filtres de Gabor [26] permettent d'obtenir des caractéristiques globales d'une image en niveaux de gris calculées par plusieurs convolutions en faisant varier l'échelle, l'orientation et la variance d'un noyau gaussien. Nous avons généré le banc de filtres d'après la publication de Zhou et Wei [65].

Les paramètres des filtres sont classiquement :

- le coefficient d'échelle : 0, 1, 2, 3 et 4
- l'orientation : 0°, 45°, 90° et 135°
- la variance de la gaussienne fixée à 0.75π et 1.5π

En utilisant ces paramètres, on obtient 40 filtres de Gabor. Chaque image est donc convoluée 40 fois, et on extrait la moyenne et l'écart type de la magnitude

(module de la réponse complexe au filtre de Gabor). La figure 4.3 illustre la réponse en magnitude aux 20 filtres correspondants à $\sigma = 1.5\pi$ pour un des décors en chevrons. Le descripteur global de chaque image est donc un vecteur de 80 réels.

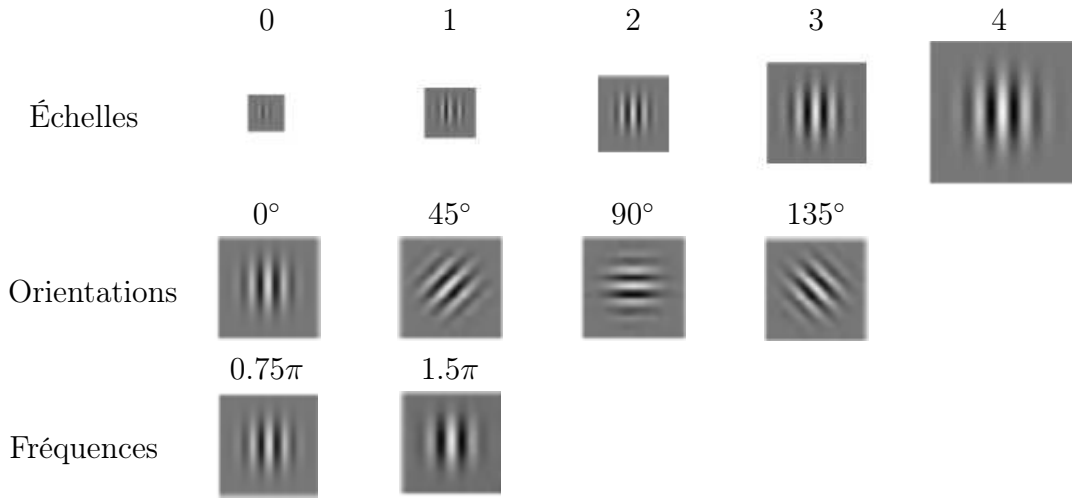


FIGURE 4.2 – Représentation des différents filtres de Gabor utilisés.

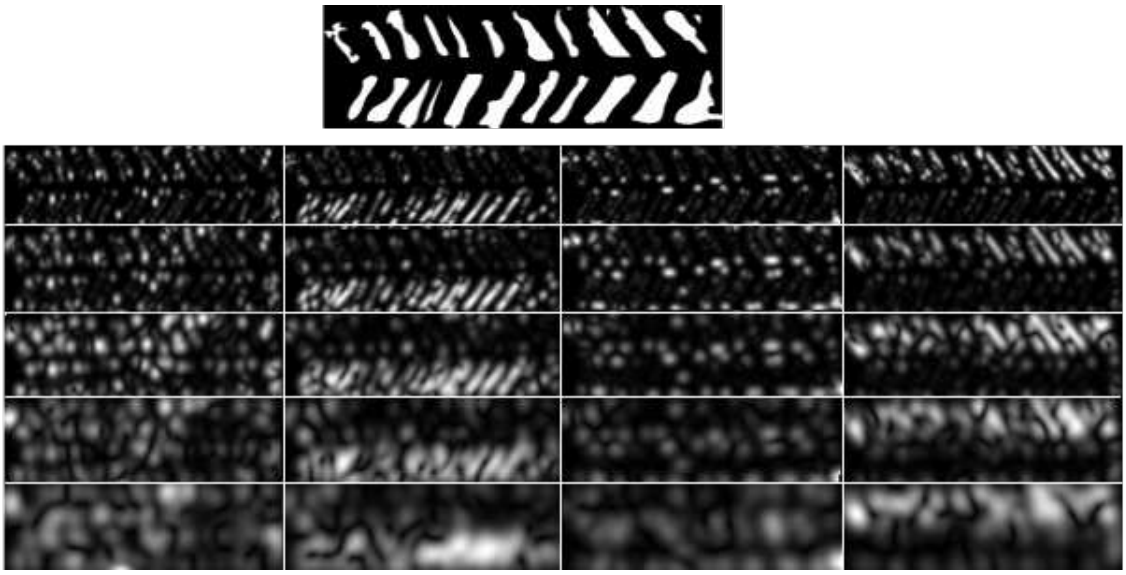


FIGURE 4.3 – Visualisation de la magnitude de la réponse pour les 20 filtres à la fréquence 1.5π (5 échelles et 4 orientations).

4.1.2 Détection de points d'intérêt et SURF

Les descripteurs locaux sont calculés autour de points d'intérêt. Un des premiers outils d'extraction automatique de points d'intérêt, proposé par Harris et

Stephen, exploite l'autocorrélation dans les images [50]. La matrice Hessienne H des dérivées partielles d'ordre 2 est utilisée pour rechercher les zones de fort changement d'intensité dans l'image. Pour les points situés dans des régions homogènes, les deux valeurs propres de la matrice Hessienne (λ_1, λ_2) sont faibles. Si une seule des valeurs propres est faible, le point est situé sur un contour, et si les deux valeurs propres sont fortes, alors le point correspond à un coin. Harris et Stephen utilisent une réponse exploitant le déterminant et la trace de la matrice Hessienne (voir Eq. 4.4).

$$R = \lambda_1\lambda_2 - k(\lambda_1 + \lambda_2)^2 = det(H) - k.trace^2(H) \tag{4.4}$$

où k règle la sensibilité du détecteur, généralement fixé à 0.02.

L'algorithme SURF (Speeded Up Robust Features) réalise une sélection de points d'intérêt dans l'image selon un principe similaire avec une approche multi-échelle [51].

Pour un point $p = (x, y)$ de l'image I, la matrice Hessienne $H(p, \sigma)$ à l'échelle σ est calculée par :

$$H(p, \sigma) = \begin{pmatrix} L_{xx}(p, \sigma) & L_{xy}(p, \sigma) \\ L_{yx}(p, \sigma) & L_{yy}(p, \sigma) \end{pmatrix} \tag{4.5}$$

où les $L_{..}(p, \sigma)$ représentent les convolutions de la dérivée seconde de l'image $I(x, y)$ au point p avec un noyau Gaussien.

Pour accélérer les calculs, SURF utilise des filtres approximatifs pour les dérivées (voir Fig. 4.4). Les changements d'échelles sont réalisés en appliquant des filtres de taille différente sur l'image initiale, au lieu de réduire itérativement la taille de l'image comme pour le détecteur Hessian-Laplacian [66](voir Eq. 4.6). La sortie du filtre 9x9 est considérée comme la première couche correspondant aux dérivées gaussiennes avec $\sigma = 1.6$, les couches suivantes sont obtenues en filtrant l'image avec des masques progressivement plus grands de taille 9x9, 15x15, 21x21, 27x27, ...

$$\sigma_{\text{approx}} = \text{taille du filtre courant} \times \frac{\text{échelle du filtre de base}}{\text{taille du filtre de base}} \tag{4.6}$$

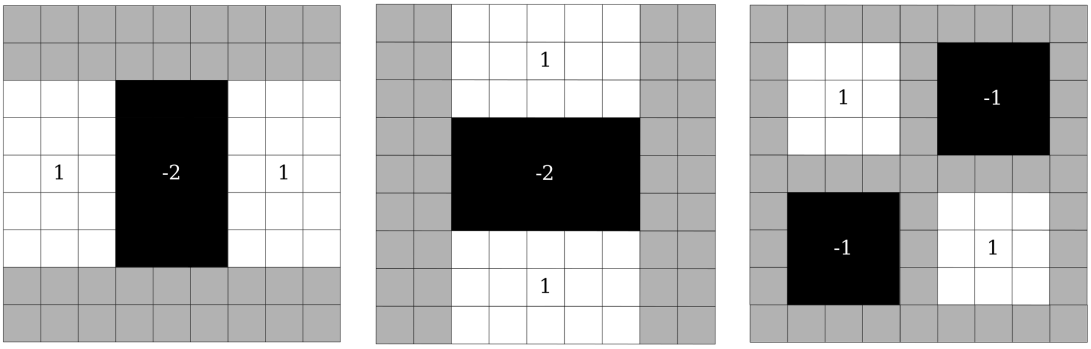


FIGURE 4.4 – Filtres de calcul des dérivées secondes.

Les points d'intérêt sont sélectionnés comme des maxima locaux dans cette représentation multi-échelles, avec une interpolation en échelle et dans l'espace image dans un voisinage 3x3x3 (voir Fig. 4.5). La figure 4.6 montre les points d'intérêt SURF détectés sur un décor binaire.

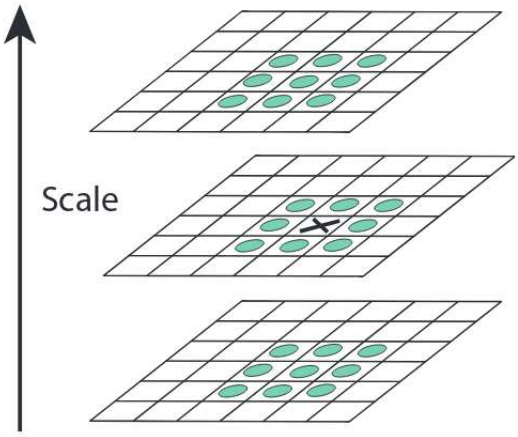


FIGURE 4.5 – Les maxima et les minima sont détectés en comparant un pixel (marqué avec X) à ses 26 voisins dans un voisinage 3x3x3 (marqué par des cercles) (source [2]).

La texture autour de ces points d'intérêt est ensuite caractérisée par des ondelettes de Haar dans un voisinage carré de taille $20s$ où s est l'échelle à laquelle le point d'intérêt a été trouvé. L'invariance en rotation du descripteur est assurée par

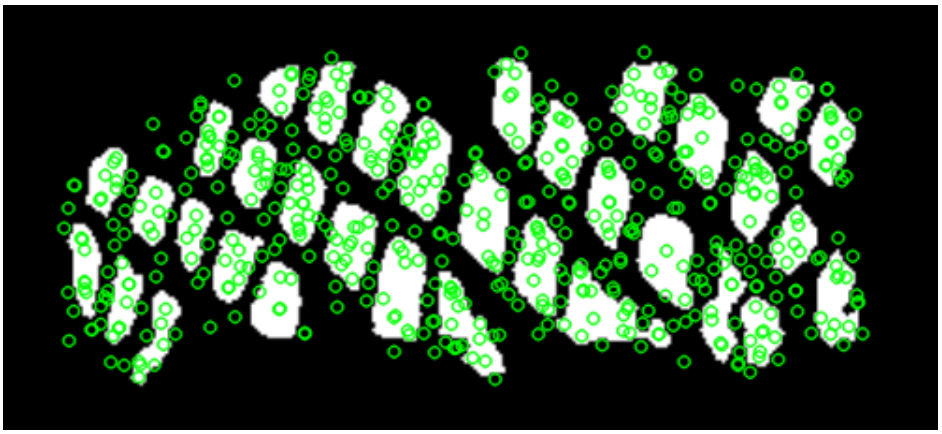


FIGURE 4.6 – Visualisation des points d'intérêt SURF.

une recherche préalable de l'orientation dominante également par les ondelettes de Haar sur un voisinage réduit de taille $6s$; le voisinage de caractérisation de taille $20s$ étant ensuite aligné sur l'orientation dominante.

Pour optimiser les temps de calcul, SURF utilise des images intégrales pour faire les calculs de convolution (voir Eq. 4.7). Le détecteur basé sur des convolutions par des filtres pour les dérivées secondes (voir Fig. 4.4) peut ainsi être calculé avec 3 opérations entières en utilisant une image intégrale pré-calculée. Son descripteur basé sur la somme des réponses en ondelettes de Haar autour du point d'intérêt peut également être calculé à l'aide de l'image intégrale. Le descripteur SURF est reconnu pour offrir un bon compromis entre robustesse et rapidité de calcul. La taille du vecteur caractéristique est de 64.

$$I_{\Sigma}(x,y) = \sum_{i=0}^{x}\sum_{j=0}^{y} I(i,j) \tag{4.7}$$

4.1.3 SIFT et SIFT Dense sur une grille régulière

En 2004, Lowe propose un descripteur particulièrement robuste au changement de point de vue [2]. D'autres détecteurs et descripteurs plus rapides ont été proposés ensuite : SURF [51], ORB [52], BRISK [53], BRIEF [67], FAST [47], FREAK [68]. Mais le descripteur SIFT reste encore aujourd'hui l'un des plus performants [69].

Le descripteur SIFT (Scale Invariant Features Transform) extrait en premier lieu les points d'intérêt comme des extrema locaux d'une différence de gaussiennes (DoG) appliquée à différentes échelles. Comme précédemment dans SURF, chaque pixel est comparé à ses 26 voisins dans le voisinage 3x3x3 de la pyramide d'images DoG (voir Fig. 4.5). Un pixel est sélectionné comme un point d'intérêt candidat s'il est un maximum ou minimum local (voir Fig. 4.7).

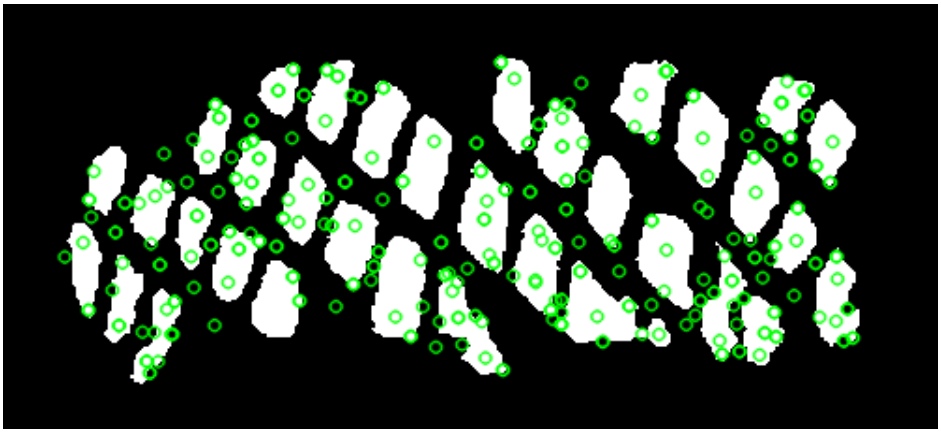


FIGURE 4.7 – Visualisation des points d'intérêt SIFT.

Chaque point d'intérêt est ensuite caractérisé par la distribution des orientations du gradient. L'invariance en rotation du descripteur est assurée par une recherche préalable des orientations dans tout le voisinage du point d'intérêt considéré. Un premier histogramme des orientations est construit par intervalles de 10 degrés pour détecter les orientations dominantes (plus de 80% de la valeur max). Puis le descripteur SIFT est construit par un histogramme des gradients (HoG) de l'intensité dans 16 quadrants de taille 4x4 autour du point considéré (voir Fig. 4.8). Les orientations sont quantifiées selon 8 valeurs. Le descripteur SIFT est de taille 4x4x8=128.

Pour le descripteur SIFT Dense [70], le détecteur n'est pas utilisé. L'extraction des descripteurs se fait sur une grille régulière espacée de M pixels. Cette approche dense donne de meilleurs résultats dans la plupart des applications de reconnaissance d'images ou d'objets [69]. Souvent, les descripteurs issus d'une grille régulière sont utilisés pour générer une représentation vectorielle à partir d'un vocabulaire visuel. La notion de vocabulaire visuel est présentée en section 4.1.4. Le pas de

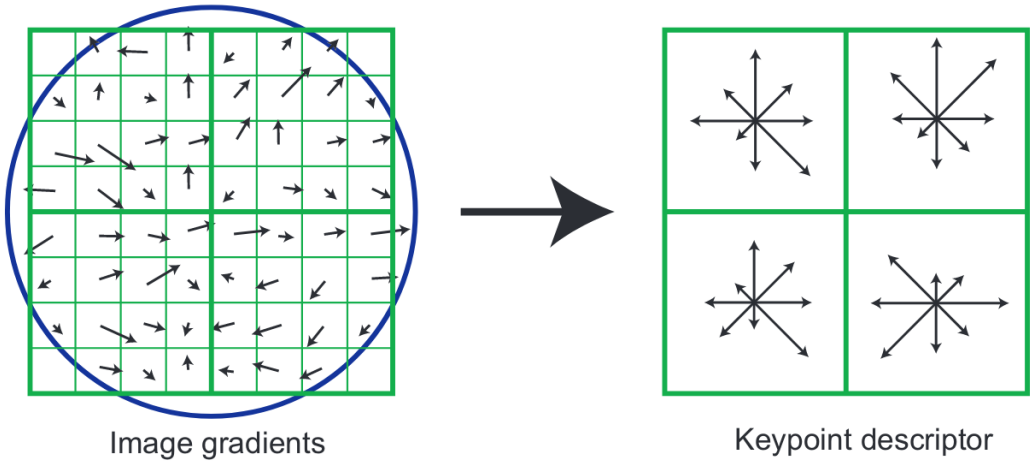


FIGURE 4.8 – Représentation du gradient de descripteurs 2x2 calculés à partir d'un ensemble d'échantillons de 8x8 (source [2]).

grille convenant le mieux à nos décors et la taille du vocabulaire visuel seront déterminés dans la section 4.3.2.

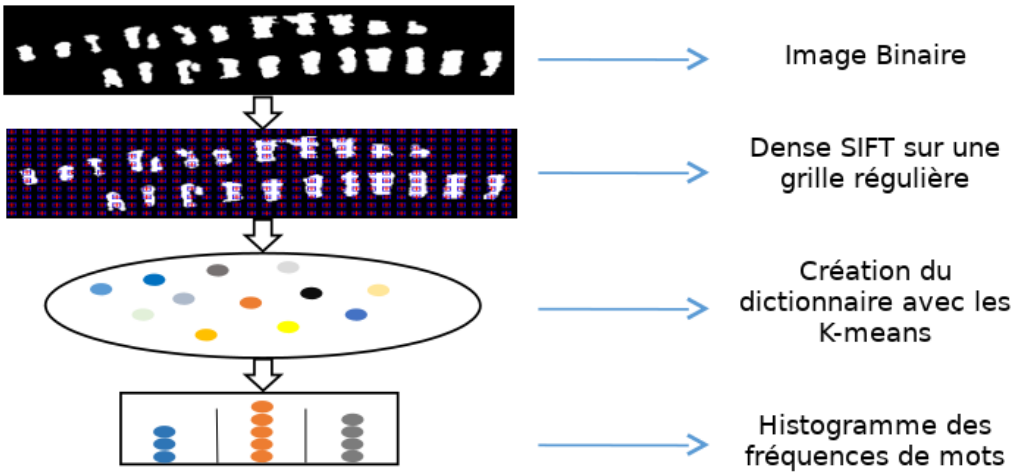


FIGURE 4.9 – Du descripteur SIFT Dense à l'histogramme de fréquence.

4.1.4 Représentation vectorielle d'une image à partir d'un vocabulaire visuel

L'exploitation directe de descripteurs locaux en vue d'une classification n'est pas toujours aisée. Dans le cas des approches à base de points d'intérêt, il est, par exemple, fréquent que le nombre de descripteurs soit variable, soit par définition, lorsque l'on recherche des extremas, soit du fait des images étudiées, qui peuvent

être de taille variable. Or les classifieurs utilisent en général, comme entrée, des vecteurs d'attributs de longueur fixe. On procède donc à une transformation permettant, pour chaque image, de synthétiser les descripteurs locaux sous la forme d'un nombre constant d'attributs. Plus précisément, nous avons retenu des représentations sous forme de vecteurs numériques, par le biais de dictionnaires de mots visuels. Rappelons qu'un tel dictionnaire est constitué d'un nombre limité de motifs –donc dans notre cas, de descripteurs locaux– représentatifs de l'ensemble des descripteurs observés (voir Fig. 4.9).

Une fois le vocabulaire établi (dictionnaire de mots visuels de taille K), l'approche classique de BoW débouche sur une représentation vectorielle de l'image sous la forme d'un histogramme en dénombrant les occurrences des mots visuels (voir Fig. 4.10), c'est-à-dire le nombre de descripteurs SIFT assigné à chaque cluster [71]. Cet histogramme de fréquence peut alors servir d'entrée à un classifieur supervisé, tel qu'un SVM.

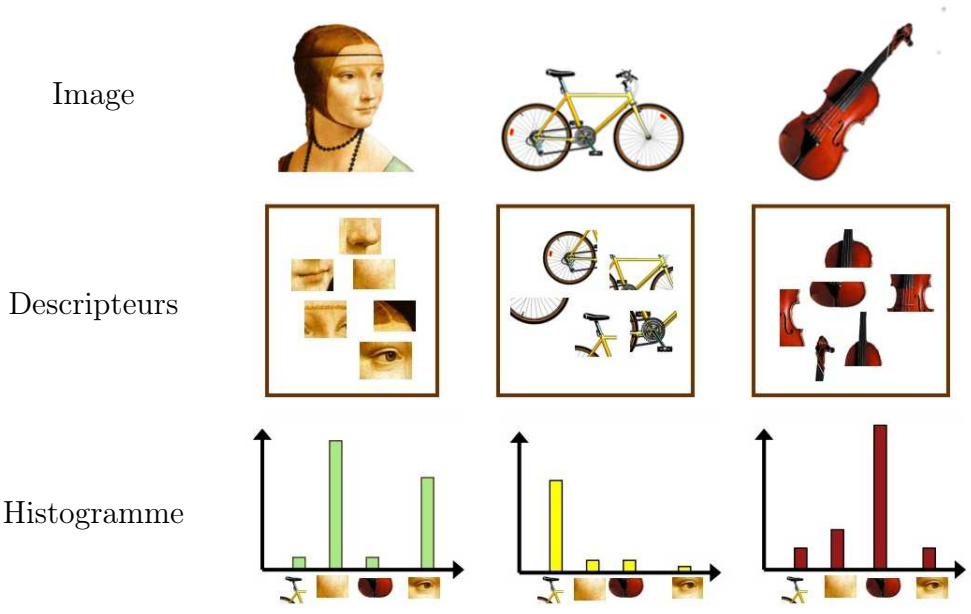


FIGURE 4.10 – Bag of Words - Représentation des images comme histogrammes d'occurrences de mots (source [3]).

Plusieurs approches ont été utilisées pour élaborer un dictionnaire des mots visuels. La première et la plus courante est la méthode d'apprentissage non supervisé (ou clustering) par K-means [58] appliquée à tous les descripteurs SIFT

extraits d'une grille régulière [71]. Le dictionnaire est paramétré par le nombre de mots souhaité K, qui correspond au nombre de centroïdes (ou clusters).

Une autre approche consiste à construire un vocabulaire visuel probabiliste basé sur un modèle de mélanges gaussiens (GMM). Rappelons que le mélange gaussien, repose sur l'hypothèse d'une distribution des objets, dans l'espace des attributs, suivant un ensemble de lois multinormales. Le modèle s'appuie alors sur K lois (K pouvant être fixé à l'avance ou estimé par l'algorithme). Les paramètres de cet ensemble de lois sont donc, d'une part, les paramètres de chaque loi normale et, d'autre part, la distribution de la probabilité qu'un objet tiré au hasard suive l'une des lois. Contrairement aux K-means, le GMM modélise un dictionnaire sous la forme de mots (ou clusters) non-disjoints, on parle alors de soft-clustering ou soft-BoW. L'algorithme itératif d'Espérance-Maximisation (EM) [72] permet d'estimer les paramètres du GMM qui s'approche le plus de la distribution du jeu de données d'apprentissage (descripteurs SIFT extraits de la grille). Comme l'algorithme EM est très sensible à l'initialisation des paramètres, l'estimation du GMM propose deux méthodes pour estimer les valeurs initiales des moyennes et des covariances : par un échantillonnage aléatoire des données (la plus rapide mais la moins robuste) ou par K-means.

Deux améliorations récentes proposent d'exploiter les résidus ou différences entre les entités locales extraites des images (descripteurs SIFT) et les mots du dictionnaire, plutôt que les fréquences des mots : les vecteurs de Fisher [73] et l'approche VLAD (Vector of Linearly Agregated Descriptors) [74]. Toutes deux permettent de limiter la taille du vocabulaire (nombre de mots du dictionnaire) et ainsi le coût de calcul.
L'approche par vecteur de Fisher [73] consiste à construire un vocabulaire visuel probabiliste basé sur un modèle de mélanges gaussiens (GMM). Le vecteur de Fisher (FV) stocke les propriétés statistiques de distribution (moyenne et covariance) de tous les modes du GMM. L'image est représentée par le gradient de la log-vraisemblance des descripteurs locaux par rapport aux paramètres du GMM.
Pour VLAD, la quantification vectorielle du dictionnaire de mots visuels repose sur les K-means à partir de tous les descripteurs SIFT collectés. Puis les résidus

(vecteurs différence) entre le descripteur SIFT (de taille 128) et les K centroïdes sont concaténés dans un unique descripteur de dimension 128xK. Il a été montré que le VLAD est une version simplifiée du FV où le GMM (soft-clustering) est remplacé par un K-means et où seul le gradient par rapport à la moyenne est considéré.
Pour le FV comme pour VLAD, les auteurs préconisent une normalisation des vecteurs pour améliorer les performance d'un classifieur linéaire.

4.1.5 Pyramide spatiale et descripteur PHOW

Le descripteur PHOW (Pyramid Histogram Of Word) [64] combine les descripteurs SIFT extraits sur une grille dense à plusieurs échelles à travers un schéma pyramidal spatial (voir Fig. 4.11). Une série de descripteurs SIFT est calculée en chaque point de la grille sur quatre supports (rayon de 4, 6, 8 et 10 pixels) pour prendre en compte les variations d'échelle. Ensuite, le regroupement par les K-means est utilisé pour construire un dictionnaire de mots visuels compilé à partir de dizaines ou de centaines de milliers de descripteurs SIFT.

A chaque niveau l de la pyramide, la grille est formée par 2^l cellules pour chaque dimension. Pour un niveau l et un nombre de mots visuels K, le descripteur PHOW est un vecteur de dimension $K \sum_{l=0}^{L} 4^l$ où L est le nombre de niveaux dans la pyramide. Dans la mise en œuvre, les auteurs ont limité le nombre de niveaux à 3 afin d'éviter un ajustement excessif [4].

Dans notre jeu de données, il n'y a pas vraiment de variations significatives d'échelle entre les décors, puisqu'ils sont gravés avec des cylindres en bois de 1,5 à 3 cm de large quelque soit le tesson. On peut donc s'attendre à ce que l'approche pyramidale de PHOW donne des performances assez similaires à ceux d'une approche basée sur un SIFT Dense appliqué sur le premier niveau.

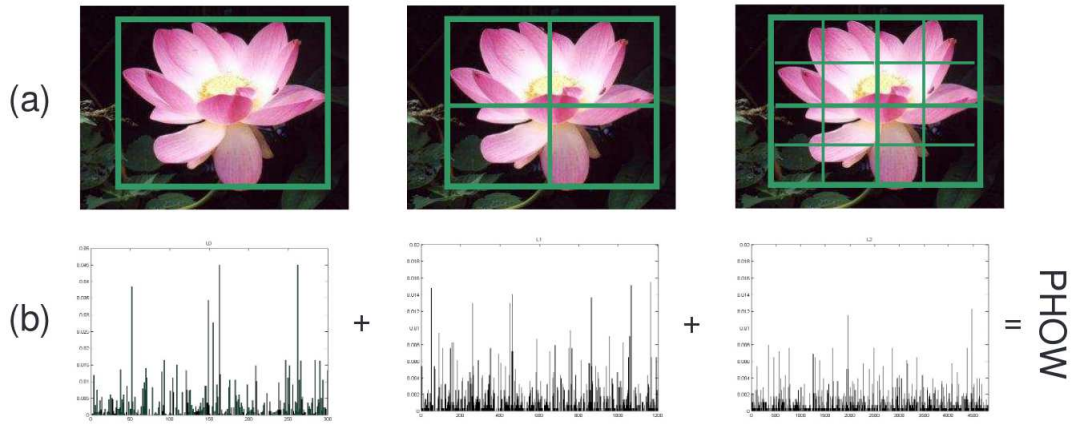


FIGURE 4.11 – (a) Représentation spatiale de la grille pour un niveau $l = 0$ à $l = 2$, (b) histogramme des représentations pour chaque niveau (source [4])

4.1.6 Approche proposée : Blob-SIFT

L'un des problèmes d'une grille régulière est que l'ensemble des descripteurs représente une grande quantité de données. Plus la taille des données est élevée plus le temps de traitement pour la classification est élevé et plus l'espace mémoire est important.
Pour la caractérisation des décors, une grande partie des points de la grille dense est située sur une zone où il n'y a pas d'information (région noire). L'idée est alors d'adapter la localisation des points d'intérêt sur les motifs périodiques du décor pour n'extraire les descripteurs SIFT que sur les zones informatives du décor.

A partir des images binaires des décors, chaque motif est isolé par une analyse en composantes connexes et localisé par son barycentre (voir Fig. 4.12). Les descripteurs SIFT sont calculés en chacun des barycentres sur trois supports de rayon $R - 8$ pixels $R - 4$ pixels et $R + 4$ pixels, où R est le rayon du cercle englobant le motif (voir Fig. 4.13). Ce descripteur est appelé Blob-SIFT car il associe la détection des motifs binaires (blob) et le descripteur SIFT. Contrairement aux descripteurs PHOW ou Dense-SIFT qui s'appliquent sur des grilles régulières, le Blob-SIFT calcule une grille adaptée à chaque décor. Cette grille adaptative présente le double avantage de collecter les signatures (descripteurs SIFT) seulement dans les zones pertinentes de motifs et de réduire considérablement la masse de données.

Le reste de l'approche BoW est standard : pour générer le dictionnaire de mots visuels, la méthode non supervisée des K-means est appliquée à tous les descripteurs SIFT extraits des motifs de la base d'apprentissage. Le dictionnaire est paramétré par K, le nombre de centroïdes qui est aussi le nombre de mots constituant le dictionnaire.

Pour classifier un décor, on extrait l'ensemble des descripteurs Blob-SIFT de l'image binaire. Chaque descripteur est ensuite associé au mot visuel du dictionnaire le plus proche, pour construire l'histogramme des occurrences de mots visuels. Le vecteur caractéristique du décor fourni en entrée au classifieur sera donc de taille K.

FIGURE 4.12 – Détection des centres et des cercles englobants de chaque motif formant le décor.

FIGURE 4.13 – Descripteurs calculés sur trois supports centrés sur chaque motif.

4.2 Classifieurs

Pour la classification, nous avons utilisé deux types de classifieurs populaires que sont les SVM et les arbres de décisions.

4.2.1 SVM

Les Séparateurs à Vaste Marge (SVM [75]) forment un modèle d'apprentissage supervisé largement utilisé dans la reconnaissance de formes pour sa capacité à travailler sur de grands ensembles de données. Rappelons que les SVM sont une approche spatiale reposant sur la recherche d'un hyperplan séparateur entre deux

classes, positionné de manière à maximiser la marge, c'est à dire la distance des objets les plus proches à cet hyperplan. Les SVM utilisent des fonctions noyaux, telles que linéaire, polynomiale, radiale et sigmoïde, pour transformer l'espace d'entité d'origine, rarement exploitable tel quel, en un espace beaucoup plus riche en dimensions. On s'attend donc à ce que la probabilité de trouver un hyperplan séparateur augmente. De plus, les noyaux présentent certaines propriétés mathématiques qui garantissent que les métriques peuvent être calculées à un faible coût en termes de complexité dans l'espace induit. Le lecteur intéressé trouvera plus d'informations sur les propriétés respectives des noyaux dans [76]. Choisir le bon noyau et le paramétrer correctement est une étape délicate qui dépend fortement de la répartition spatiale des données, ce qui est généralement difficile à percevoir dans une grande dimensionnalité. Nous avons utilisé la bibliothèque LibSVM [77] pour la classification SVM et la validation croisée pour estimer les différents paramètres.

Par ailleurs, les SVM ont été initialement conçus pour la classification binaire, puis étendus pour la classification multi-classes selon deux stratégies : "un-contre-tous" et "un-contre-un". Considérons un problème de classification comportant K classes. L'approche "un-contre-tous" consiste à construire K SVM, chacun séparant les objets d'une classe donnée contre ceux de toutes les autres classes. Tout nouvel objet sera alors classé par les K SVM, et sera considéré comme appartenant à la classe pour laquelle il est le plus éloigné de l'hyperplan, et se trouvant donc le plus loin du demi-espace des "autres classes". L'approche "un-contre-un" va pour sa part reposer sur $K\text{x}(K-1)$ SVM, chacun opposant deux classes données. Tout nouvel objet sera testé dans les $K\text{x}(K-1)$ SVM, et affecté à la classe dans laquelle il a le plus souvent été classé. Dans nos travaux, nous avons utilisé la stratégie "un-contre-un" avec un classifieur de type C-SVM.

La section 4.3.1 présente une comparaison des différents noyaux sur notre base de données avec le descripteur SIFT Dense et BoW.

4.2.2 Arbre de décision

Les arbres binaires de décision (CART - Classification And Regression Trees) [78] font partie des classifieurs supervisés classiques. Ils reposent sur un enchaînement arborescent de conditions posées sur les valeurs des attributs (ou descripteurs des objets considérés). Chaque nœud interne de l'arbre indique un descripteur cible et chaque branche issue de ce nœud une valeur (symbolique) ou plage de valeurs (numérique) pour ce descripteur. Le parcours se fait ainsi jusqu'à atteindre une feuille, laquelle contient une étiquette (ou label) de classe. Différents mécanismes d'apprentissage sont disponibles, les plus efficaces s'achevant par une phase d'élagage, ou pruning, supprimant certaines branches dans le but de limiter le phénomène de surapprentissage, donc afin de rendre l'arbre plus robuste face à de nouveaux objets.

La combinaison ou l'agrégation de modèles permet généralement d'améliorer les performances de l'apprentissage automatique [79], notamment le problème du sur-ajustement des modèles. Deux types de stratégies ont été abordés : aléatoires (bagging [80], random forest [81]) ou adaptatives (boosting [82], gradient boosting). Les principes du bagging ou du boosting s'appliquent à toute méthode de modélisation (régression, arbres de décision, réseaux de neurones).

Le principe du bagging est élémentaire : il consiste à combiner les réponses de plusieurs modèles par une simple moyenne pour les variables quantitatives, ou faire voter un comité de modèles pour déterminer la réponse la plus probable dans le cas d'une variable qualitative, cela permet de réduire la variance et donc l'erreur de prévision. Les modèles sont entrainés sur des réplications d'échantillons (bootstrap) obtenues par des tirages aléatoires avec remise dans le jeu de données.

Dans le cas spécifique des forêts aléatoires, Breiman [81] propose une amélioration du bagging afin de rendre les arbres plus indépendants. Sur chaque tirage aléatoire d'échantillons (bootstrap), un arbre est estimé avec randomisation des variables : la recherche de chaque division optimale est précédée d'un tirage aléatoire des variables.

Dans la famille des stratégies adaptatives, l'algorithme adaBoost (Adaptative boosting) a été décrit à l'origine pour la prévision d'une variable binaire [82], puis généralisé à K classes. Dans le boosting, la construction des modèles, qui seront ensuite agrégés par une moyenne pondérée, est faite de façon récurrente : chaque modèle est une version adaptée du précédent en se concentrant sur les observations ayant obtenues les plus mauvaises prédictions. Les nombreuses variantes proposées dans la littérature différent par le type de variables à prédire (binaire/K classes, réelles/qualitatives), la façon de renforcer l'importance des observations mal estimées, la façon d'agréger les modèles. Le dernier proposé dans la librairie XGBoost (extrem gradient boosting) connaît un grand succès [83]. Une nouvelle fonction objectif est considérée en complétant la fonction perte convexe différentiable de l'apprentissage par un terme de régularisation pour éviter le sur-ajustement. L'approximation du gradient de la fonction objectif par un développement de Taylor au second ordre permet une parallélisation efficace de la construction des arbres : il suffit alors de sommer, pour chaque feuille, les valeurs des dérivées première et seconde de la fonction perte.

L'algorithme original adaBoost ne nécessitait pas de réglage fin de paramètres mais la nouvelle version (extrem gradient boosting) plus performante a multiplié le nombre de paramètres à régler et à optimiser.

Pour nos tests, nous avons utilisé la librairie XGBoost. Les arbres ont une profondeur maximum de 6 avec pour fonction objectif Softmax pour une classification multi-classes.

4.3 Expérimentations

Les sections suivantes présentent les tests réalisés pour choisir le meilleur paramétrage pour le descripteur SIFT Dense + BoW et la source d'informations la plus pertinente : carte de profondeurs, carte des variances ou image binaire. Enfin une comparaison des différentes approches est proposée.

4.3.1 Comparaison des noyaux SVM pour SIFT Dense et BoW

Les différents noyaux comparés pour le classifieur SVM utilisent les mêmes entrées, c'est-à-dire l'approche SIFT Dense+BoW avec un dicitionnaire de 200 mots visuels extrait sur les images binaires de la base 1 (377 images). Les tests pour fixer la taille du dictionnaire seront présentés dans la section 4.3.2. Les tests de classification effectués sur notre ensemble de données avec différents noyaux sont donnés dans la table 4.1. Pour chaque noyau, 20 tirages aléatoires du jeu de données sans recouvrement sont effectués, avec la répartition : 66% pour l'apprentissage, 33% pour le test. La table 4.1 donne les taux de reconnaissance moyens pour quatre noyaux : la fonction de base radiale (RBF), le noyau χ^2, la régression linéaire et le noyau d'intersection de l'histogramme (Inter). Le noyau χ^2 donne le meilleur taux sur notre ensemble de données avec 84.76% pour un paramètre d'échelle γ fixé à 1. Son efficacité a d'ailleurs été démontrée dans plusieurs études antérieures [84, 85] pour la classification des descripteurs issus des BoW et pour la classification multi-classes.

TABLE 4.1 – Taux moyens de bonne classification SVM des différents noyaux sur 20 tirages aléatoires sur l'ensemble des données de la base 1 (66% pour l'apprentissage, 33% pour le test).

Noyau du SVM		**Moyenne (%)**
χ^2	$K(x,y) = e^{-\gamma^2(x,y)}$	84.76 ± 6.81
Inter	$K(x,y) = \sum_i min(x_i, y_i)$	83.81 ± 6.97
Linéaire	$K(x,y) = x^T y$	81.90 ± 9.05
RBF	$K(x,y) = e^{-\gamma\|\|x,y\|\|^2}$	81.75 ± 9.19

4.3.2 Paramétrage du descripteur SIFT Dense + BoW

Comme présenté précédemment, BoW consiste à construire un dictionnaire de K mots visuels à partir de descripteurs SIFT Dense calculés sur une grille régu-

lière avec un espacement de M pixels à plusieurs échelles (sur quatre supports circulaires différents de rayon de 4, 6, 8 et 10 pixels). Ce qui fournit un vecteur de taille 4×128 pour chaque point. Pour définir les paramètres, les taux moyens de bonne classification ont été observé en faisant varier le vocabulaire de 20 à 270 mots visuels et l'espacement de la grille de 8 à 16 pixels (avec un pas de 10 mots et 2 pixels, respectivement). Le noyau adopté pour SVM est le noyau gaussien généralisé avec une distance χ^2. Dix tirages aléatoires du jeu de données sans recouvrement ont été effectués pour chaque combinaison des paramètres en divisant l'ensemble de données : 66% pour l'apprentissage, 33% pour le test sur la base 1 (377 images). D'après la figure 4.14, on constate qu'un dictionnaire d'une quarantaine de mots permet déjà d'atteindre un taux moyen de bonne classification de 70%. La progression du taux moyen de bonne classification se ralentit ensuite. La meilleure combinaison est un dictionnaire de 200 mots généré à partir d'une grille espacée de 12 pixels. Ces paramètres seront gardés pour la suite.

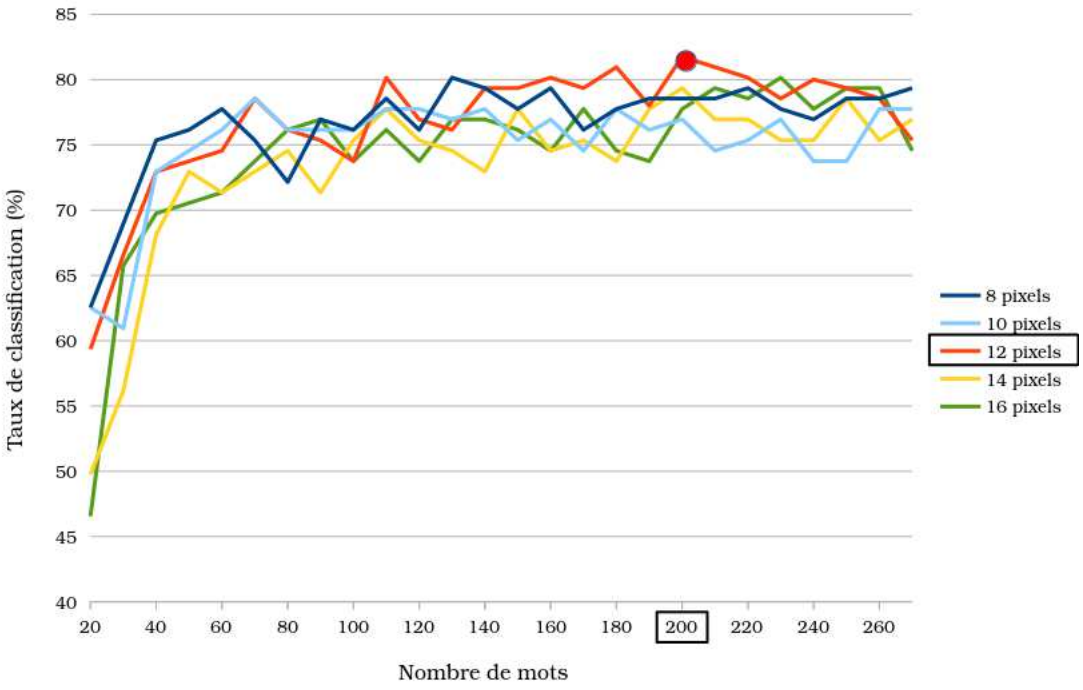


FIGURE 4.14 – Taux moyens de bonne classification sur 10 essais aléatoires sans recouvrement sur l'ensemble des données (66% pour l'apprentissage, 33% pour le test sur la base 1) en faisant varier le nombre de mots visuels et l'espacement de la grille régulière.

4.3.3 Comparaison des différentes sources d'images

Le décor du tesson peut être représenté par trois types d'images différentes : la carte de profondeur, la carte des variances locales et l'image binaire. Nous avons testé l'approche Blob-SIFT et le classifieur SVM (noyau χ^2) sur ces 3 types d'images sur la base 1 (377 images). Le descripteur SIFT initialement conçu pour des images en niveaux de gris telles que la carte de profondeurs ou des variances, peut s'appliquer directement aux images binaires. Pour les trois différents types d'entrées, la même région de saillance est prise en compte. Les coordonnées des barycentres des motifs sont calculées sur les images binaires. Les mêmes coordonnées sont utilisées pour positionner le descripteur SIFT sur les trois types d'images. La classification est faite sur 20 tirages aléatoires du jeu de données sans recouvrement, avec la répartition : 66% pour l'apprentissage, 33% pour le test. Les résultats sont indiqués dans la table 4.2.

TABLE 4.2 – Comparaison des taux moyens de bonne classification avec l'approche Blob-SIFT sur les différentes sources d'images : moyenne et variance pour 20 tirages aléatoires sans recouvrement sur la base 1 (66% pour l'apprentissage, 33% pour le test).

Entrées du SVM multi-classes	**Moyenne (%)**
Cartes de profondeurs	72.30 ± 10.27
Cartes des variances locales	80.00 ± 13.76
Images binaire	84.76 ± 6.81

La classification à partir des images binaires surpasse celle des cartes de variances d'environ 5% et celle des cartes de profondeurs d'environ 10%. Dans le même sens, les variances sur les 20 tirages aléatoires dénotent bien une meilleure stabilité des taux de classification pour les images binaires. Pour la suite, les images binaires seront utilisées.

4.3.4 Pertinence de l'étape d'extraction de la région de saillance

La binarisation directe de la carte de profondeurs de l'ensemble du tesson introduit de nombreux artefacts dus à la texture granuleuse de l'argile (voir Fig. 4.15). L'extraction de la région de saillance centrée sur le décor vise à être plus proche de l'encrage manuel réalisé par l'archéologue. Nous avons évalué l'influence de cette étape d'extraction sur les taux moyens de bonne classification avec l'approche développée et le classifieur SVM (noyau χ^2). Les matrices de confusion avec et sans l'étape d'extraction sont présentées dans les tables 4.3 et 4.4. Les tables donnent le pourcentage de la classe prédite par rapport à la classe réelle. Pour rappel, classe A : losanges, classe C-G : bâtons ou carrés sur deux registres, classe H : carrés sur trois registres, classe L : chevrons.

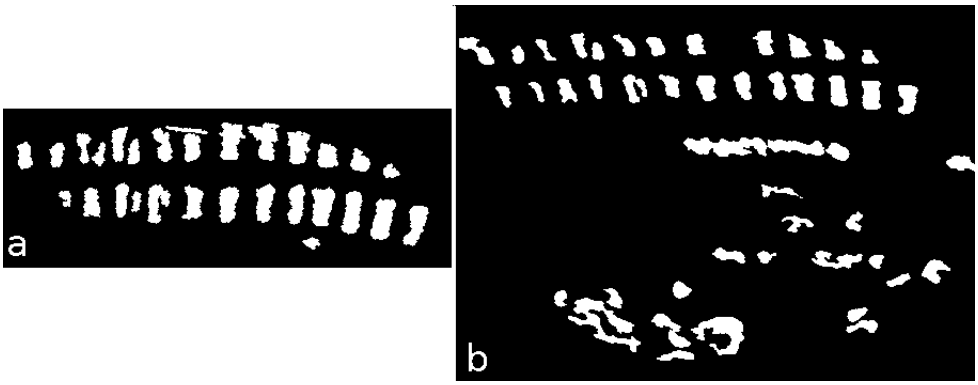


FIGURE 4.15 – Binarisation de la carte de profondeurs avec (a) et sans (b) application de l'étape d'extraction de la région saillante.

Avec l'étape d'extraction (voir table 4.4), les taux moyens de bonne classification des types A, C-G et H sont supérieurs à 80% alors que ceux des types B et L sont plus faibles. Le type B est le motif le plus variable, il n'est pas aussi répétitif que les autres types, et c'est la classe la moins abondante dans l'ensemble de données. Le type L est également sous-représenté dans l'ensemble de données. Ainsi, l'étape d'apprentissage n'est probablement pas suffisante pour une classification correcte de ces deux classes.

Si l'on compare les pourcentages dans les tables 4.3 et 4.4, l'étape d'extraction est très pertinente puisqu'elle augmente le taux moyen de bonne classification

TABLE 4.3 – Matrice de confusion sans l'étape d'extraction de la région saillante. Taux moyen de bonne classification : 65.60% (var. 10.18%).

		Classes Prédites				
Classes Réelles	**%**	**A**	**B**	**C-G**	**H**	**L**
	A	**59.88**	1.81	19.95	16.78	1.58
	B	17.26	**10.13**	56.55	5.35	10.71
	C-G	6.59	1.10	**66.91**	24.30	1.10
	H	5.95	0.10	12.60	**81.25**	0.10
	L	17.46	7.94	31.75	3.70	**39.15**

TABLE 4.4 – Matrice de confusion avec l'étape d'extraction de la région saillante. Taux moyen de bonne classification : 84,76% (var. 6,81%).

		Classes Prédites				
Classes Réelles	**%**	**A**	**B**	**C-G**	**H**	**L**
	A	**90.48**	1.36	6.35	1.36	0.45
	B	18.45	**53.57**	16.67	1.19	10.12
	C-G	1.71	0.49	**83.88**	11.97	1.95
	H	0.50	0.20	8.43	**90.87**	0.00
	L	6.88	4.76	17.46	0.53	**70.37**

d'environ 20% (84,76% avec étape d'extraction contre 65,60% sans) et réduit les confusions entre certaines classes.

La figure 4.16 rappelle les différents décors des classes. La figure 4.17 donne trois exemples de décors mal classés. La première confusion entre les types C-G (bâtons-carrés sur 2 registres) et H (carrés sur 3 registres) peut s'expliquer par certains artefacts qui n'ont pas été enlevés lors du prétraitement (points rouges). La confusion entre les types L et C-G est due à l'inclinaison insuffisante des chevrons qui les fait ressembler à des bâtons. Pour la confusion entre les types A et B, on remarque un manque de motifs qui rend le décor non répétitif. Il est donc classé comme type B.

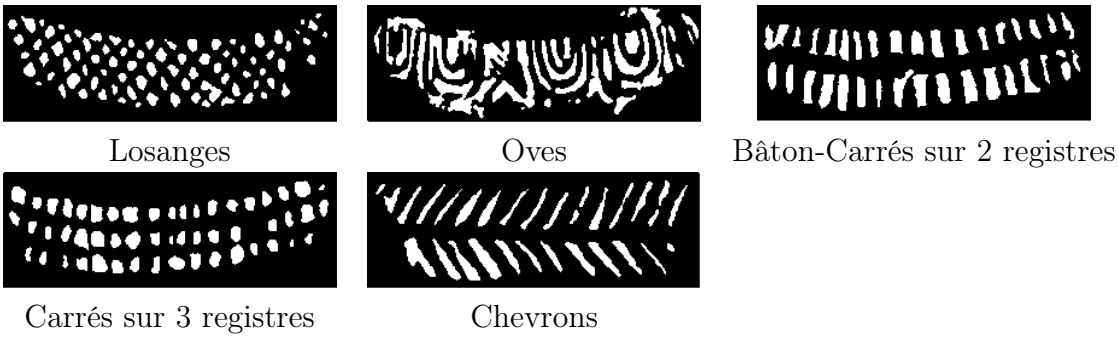


FIGURE 4.16 – Rappel des différentes représentations des décors (empreintes manuelles).

FIGURE 4.17 – Décors mal classés.

4.3.5 Comparaisons de différentes approches

Dans cette section, nous comparons les taux moyens de bonne classification obtenus avec l'approche proposée Blob-SIFT et les principaux descripteurs populaires de la littérature. Pour cette comparaison, nous avons retenu une approche globale par les filtres de Gabor et plusieurs approches locales par sac-de-mots visuels. Les deux premières approches locales exploitent respectivement les détecteurs et descripteurs SURF [51] et SIFT [2], avec un dictionnaire de 200 mots obtenu par K-means. Les quatre approches suivantes utilisent des descripteurs SIFT Dense (D-SIFT) calculés sur une grille régulière avec un espacement de 12 pixels sur quatre supports circulaires (4, 6, 8 et 10 pixels). Ce qui fournit un vecteur de taille 128x4 pour chaque point de la grille.

C'est dans l'étape de construction de mots visuels à partir de ces descripteurs que diffèrent ces approches locales. D-SIFT [66] utilise une approche de BoW classique : le dictionnaire de 200 mots et l'occurence des mots visuels exploitent

K-means. Dans le cas de D-SIFT + normL2 [67], une étape de normalisation par la norme L2 est ajoutée. Les variantes D-SIFT + Fisher [69] et D-SIFT + VLAD [70] reposent respectivement sur le vecteur de Fisher issu d'un GMM pour la première, et le vecteur de résidus et l'approche K-means pour la seconde, avec une taille de dictionnaire réduite à 64 mots. Enfin, PHOW [60] met en oeuvre une structure pyramidale pour extraire les descripteurs SIFT Dense sur 3 niveaux d'échelle, puis une approche classique de BoW avec un dictionnaire de 200 mots. Les implémentations de VLAD, Fisher et PHOW sont basées sur la bibliothèque VLFeat [86]. Les autres approches sont basées sur la bibliothèque OpenCV [87].

La table 4.6 donne pour ces neuf approches les performances obtenues pour les deux classifieurs testés : SVM et XGBoost. Les pourcentages de bonne classification affichés représentent la moyenne et la variance obtenus sur 20 tirages aléatoires sans recouvrement de la base 1 (377 images binaires), avec 66% pour l'apprentissage et 33% pour les tests, en respectant la proportion des cinq classes.
Pour le classifieur SVM, les tests ont été réalisés avec le noyau le mieux adapté à chaque approche. Comme on l'a vu dans la section 4.3.1, c'est le noyau χ^2 qui est le mieux adapté à nos données pour l'approche classique de BoW avec les descripteurs SIFT. Dans le cas du vecteur de Fisher et de VLAD, c'est le noyau linéaire qui est préconisé.
La colonne « Taille des données » correspond à l'espace mémoire occupé pour le stockage des descripteurs après leur extraction dans les 377 images de la base d'apprentissage. On peut avoir une idée de la complexité des descripteurs exploités par chaque approche par la taille moyenne des données extraites par image (en divisant cet espace mémoire par 377). La taille des vecteurs transmis au classifieur est fixée par la taille du vocabulaire visuel : 64 mots pour Fisher et VLAD et 200 pour toutes les autres (voir table 4.5).

4.4 Discussion

Pour les deux classifieurs, les tentatives de caractérisation globale (Gabor) sont moins efficaces que l'utilisation de descripteurs locaux autour de points d'intérêt. Les descripteurs SIFT et SURF présentent des performances proches, les diffé-

TABLE 4.5 – Comparaison des différentes approches testées.

Approches	**Descripteur**	**Dictionnaire**	**Classifieur**
Gabor [26]	Filtres de Gabor (taille 80)	-	SVM χ^2
SURF [51]	Points d'intéret et descripteur SURF	K-means (200 mots)	SVM χ^2
SIFT [2]	Points d'intéret et descripteur SIFT	K-means (200 mots)	SVM χ^2
D-SIFT [70]	SIFT sur une grille (12 pixels)	K-means (200 mots)	SVM χ^2
D-SIFT + norme L2 [71]	SIFT sur une grille (12 pixels) NormL2	K-means (200 mots)	SVM χ^2
D-SIFT + Fisher [73]	SIFT sur une grille (12 pixels)	GMM (64 mots) + Vecteurs de Fisher (128x64x2)	SVM-Linéaire
D-SIFT + VLAD [74]	SIFT sur une grille (12 pixels)	VLAD + K-means (64 mots)	SVM-Linéaire
PHOW [64]	D-SIFT sur pyramide (3 niveaux)	K-means (200 mots)	SVM χ^2
Blob-SIFT	SIFT centré sur les motifs (grille adaptative)	K-means (200 mots)	SVM χ^2

rences peuvent en partie s'expliquer par une localisation des points d'intérêt plus dense pour SURF. L'extraction des descripteurs SIFT sur une grille régulière améliore le taux de 3% à 5% comparé à l'utilisation du détecteur intégré à SIFT selon le classifieur. La normalisation L2 n'apporte pas d'amélioration significative. L'approche pyramidale du descripteur PHOW augmente considérablement la quantité de données extraites des images sans améliorer les taux de reconnaissance. Comme on pouvait s'y attendre, nos décors ne présentant pas de variation d'échelle significative, PHOW n'apporte pas d'information pertinente supplémentaire par rapport à une unique grille dense et ses performances sont similaires à celle d'un SIFT-Dense.

Malgré sa complexité, la performance du noyau de Fisher et du GMM est décevante. L'approche VLAD par les résidus est d'un niveau de performance identique

TABLE 4.6 – Taux moyens de bonne classification des différentes approches pour 20 tirages aléatoires sur la base 1 (66% pour l'apprentissage, 33% pour le test).

Approches	**Taille des données (Mo)**	**SVM (%)**	**XGBoost (%)**
Gabor	0.273	61.04 ± 19.6	56.38 ± 11.3
SURF	77.5	80 ± 11.46	72.36 ± 14.25
SIFT	29.9	78.13 ± 13.91	71.08 ± 12.30
D-SIFT	86.6	80.95 ± 3.62	76.72 ± 9.27
D-SIFT + norme L2	86.6	81.22 ± 11.50	76.42 ± 10.73
D-SIFT+Fisher	86.6	72.03 ± 18.72	73.90 ± 19.29
D-SIFT+VLAD	86.6	79.09 ± 3.99	74.63 ± 5.45
PHOW	464.3	80.07 ± 8.17	76.19 ± 8.50
Blob-SIFT	9.2	84.76 ± 6.81	80.38 ± 12

à SIFT Dense avec un BoW classique pour SVM, un peu en-deçà pour XGBoost. L'approche proposée Blob-SIFT donne le meilleur résultat de bonne classification avec un taux de 84.76%. En se focalisant sur les motifs, Blob-SIFT permet un gain d'environ 5% sur les meilleures approches de l'état de l'art, avec une réduction drastique des données extraites des images.

Afin de confirmer ces résultats, l'approche proposée a été testée sur la base 2 (888 tessons) avec les deux classifieurs SVM et XGBoost. Les taux de classification sont fournis dans la table 4.7. Le comportement de l'approche est confirmé avec un taux de reconnaissance de 84.14% pour SVM, surpassant XGBoost d'environ 2.5%. On peut noter également une forte baisse de la variance des taux selon les tirages, expliquée par l'augmentation de la base.

La matrice de confusion de l'approche proposée Blob-SIFT avec SVM est donnée dans la table 4.8.

Nous remarquons que la majeure partie des confusions se fait entre les décors sur deux ou trois registres. Le fait d'équilibrer les classes, contrairement à la base 1, permet une meilleure séparation de la classe L. Comme nous pouvons le constater,

TABLE 4.7 – Taux moyens de classification avec l'approche Blob-SIFT sur la base 1 et la base 2.

	SVM (%)	**XGBoost (%)**
Base 1	84.76 ± 6.81	80.38 ± 12
Base 2	84.14 ± 2.93	81.73 ± 3.10

TABLE 4.8 – Matrice de confusion de la classification SVM (noyau χ^2) avec l'approche Blob-SIFT sur la base 2.

		Classes Prédites			
	%	**A**	**C-G**	**H**	**L**
Classes Réelles	**A**	**92.62**	2.01	1.21	4.16
	C-G	5.19	**77.38**	15.10	2.33
	H	3.33	12.65	**83.46**	0.56
	L	6.94	7.64	0.10	**85.32**

il y a une confusion qui pourrait surprendre entre les classe A et L. En effet, les chevrons sont des motifs éloignés des losanges. Pourtant, sur certains tessons, cette confusion s'explique par une binarisation imparfaite causée par l'état de conservation du tesson. Sur les trois premières cartes de profondeurs de la figure 4.18, le décor est sur le bord du tesson. Ces décors sont donc incomplets. Lors de la binarisation, seule une ou deux rangées de losanges apparaissent. Le décor des images binaires associé peut être interprété comme des chevrons. Pour la dernière carte de profondeurs, le décor a quasiment disparu. Le décor de l'image binaire ressemble d'avantage à des losanges qu'à des chevrons.

Un point important à souligner est que les tessons n'ont pas été particulièrement choisis lors de la création de la base 2 si ce n'est un soin à équilibrer plus équitablement les classes. Sur certains tessons, il est difficile de définir le décor (voir Fig. 4.19). Ces tessons ont été volontairement considérés pour évaluer les limites des différents algorithmes.

Un autre problème intervient dans la phase de classification. Pour les décors sur

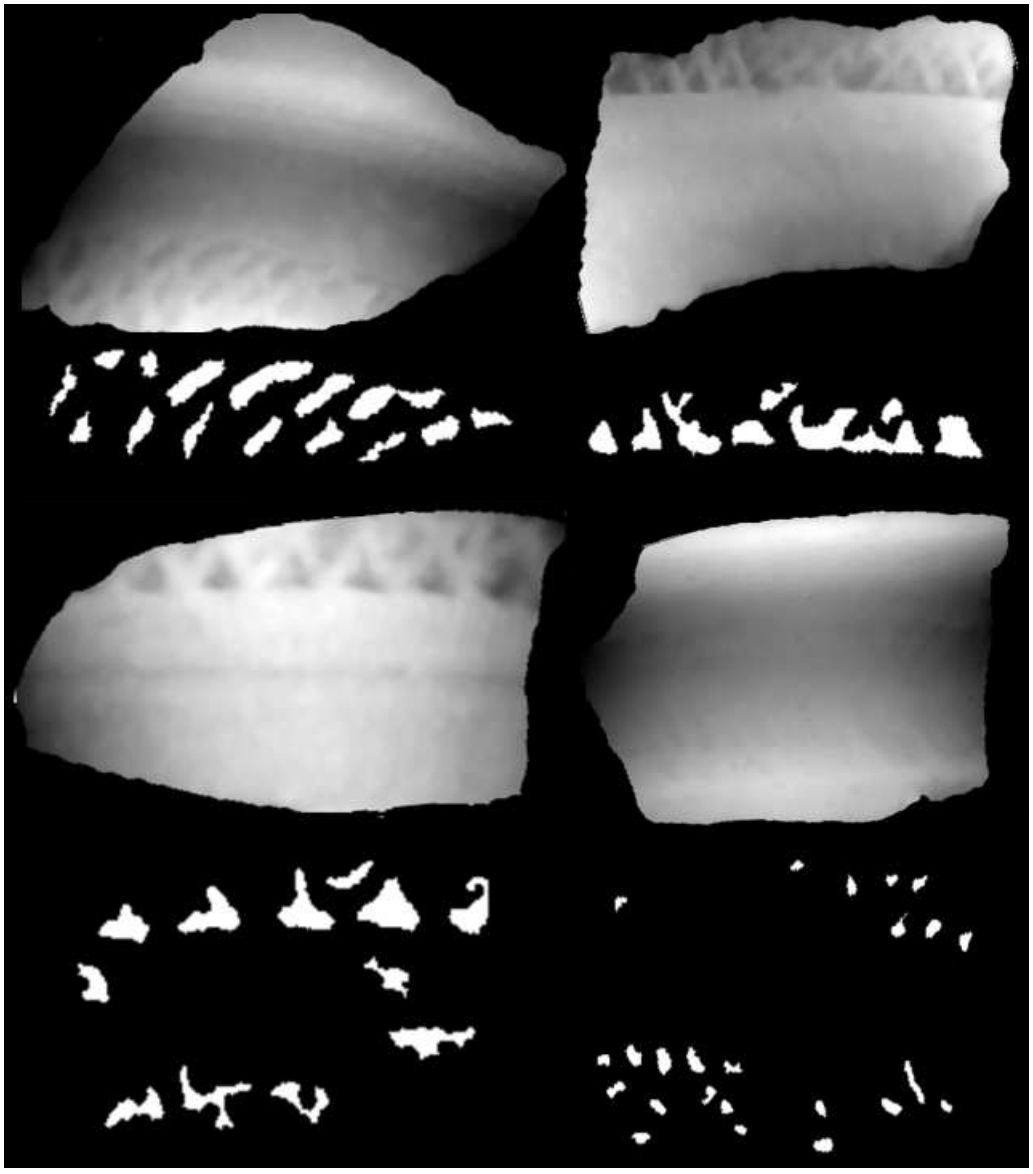


FIGURE 4.18 – Décors mal classés.

trois registres, il arrive que le troisième registre se situe sur le bord du tesson. Les motifs sont donc incomplets. Lors de la binarisation, ces demi-motifs sont considérés comme du bruit et sont donc supprimés. Le décor n'a plus que deux registres (voir Fig. 4.20). C'est l'une des raisons pour lequel il y a autant de confusions entre les classes C-G et H. Ces tessons problématiques représentent 20% de la base 2. La classification sans ces tessons augmente de 4%.

Dans ce chapitre nous avons comparé différentes approches basées sur des descripteurs globaux et locaux extraits sur les trois modalités d'images (variance, profondeur et binaire). Nous avons réalisé différents tests pour optimiser et fixer

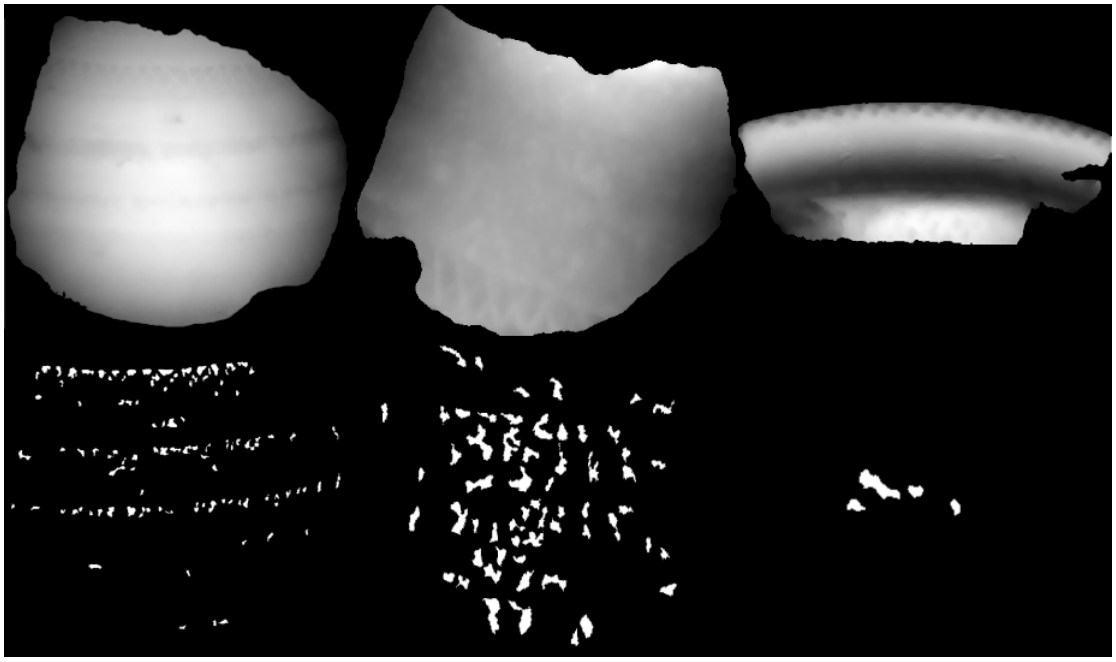


FIGURE 4.19 – Tessons dont il est difficile d'extraire le décor.

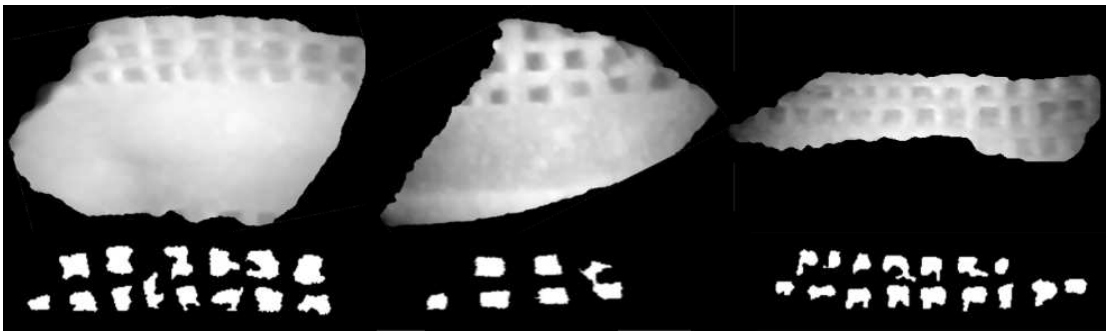


FIGURE 4.20 – Décors binaires sur 2 registres alors que les décors sont sur 3 registres sur les cartes de profondeurs.

les paramètres. Ces étapes nous ont permis de proposer une approche pertinente exploitant l'information binaire. Dans le chapitre suivant, nous allons confronter les résultats obtenus aux approches d'apprentissage profond (Deep learning).

Chapitre 5

Apprentissage profond

Ces dernières années, le développement des capacités de calcul a permit l'émergence des techniques d'apprentissage profond (deep learning). Ces techniques ont rapidement suscité un engouement qui s'est traduit par la proposition d'une multitude de méthodes efficaces qui représentent l'état de l'art [88] dans plusieurs domaines (détection d'objets [89], segmentation [90], biométrie [91]). Dans ce chapitre, nous proposons de comparer les performances obtenues par l'approche développée (voir section 4.1.6) aux approches d'apprentissage profond et notamment celles basées sur les réseaux de neurones convolutionnels (CNN : Convolutional Neural Network). Les résultats présentés dans le chapitre précédent seront comparés à différents modèles CNN et une proposition de fusion des deux approches sera proposée.

Avant de décrire les modèles utilisés, une brève introduction des modèles CNN est présentée dans la section suivante.

5.1 Réseaux de neurones convolutifs : introduction

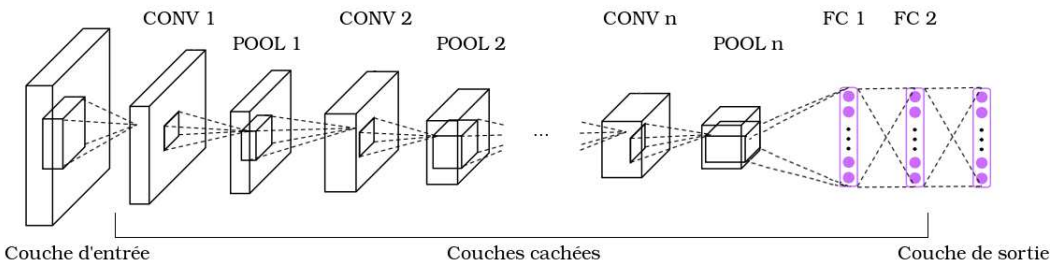


FIGURE 5.1 – Exemple d'un réseau de neurones convolutionnel.

Les CNN sont maintenant l'approche dominante pour une grande partie des

tâches de reconnaissance et de détection [92, 93] avec des résultats proche des capacités humaine. Ce qui a amené la plupart des grandes entreprises technologiques, dont Google [94], Facebook [95], Microsoft [96], Yahoo ! [97] et Adobe[98], ainsi qu'un nombre croissant de start-ups à lancer des projets de R&D tout en déployant des produits et services de compréhension d'images basés sur les réseaux de neurones convolutifs.

Egalement appelé ConvNet ou CNN, ce type de réseau est généralement constitué d'une succession d'opérations : convolution (CONV) (voir Fig. 5.2), sous-échantillonnage (POOL : pooling), fonction d'activation et de neurones interconnectés (FC : Fully Connected). Ce type de réseau est souvent constitué d'une couche d'entrées (image ou patches), de plusieurs couches cachées et d'une couche de sortie (voir Fig. 5.1). En fonction de l'application, l'entrée principale est un ensemble d'images [99] et la sortie peut être une image, un scalaire ou un vecteur. Les couches cachées ont pour rôle de modifier l'entrée d'une manière non linéaire pour que les classes deviennent linéairement séparables dans la dernière couche. Une couche est définie comme un ensemble d'unités de traitement (neurones).

Les caractéristiques apprises dans la première couche d'un réseau représentent les contours pour des orientations et emplacements définis de l'image d'entrée (souvent assimilé aux filtres de Gabor [26]). La deuxième couche détecte des motifs par des arrangements particuliers d'arêtes détectées dans la couche précédente. La troisième couche associe des motifs qui correspondent à une même famille (voir Fig. 5.3). Pour les tâches de classification, les couches supérieures du réseaux amplifient les caractéristiques de l'entrée qui sont importantes pour la discrimination et suppriment les différences non pertinentes.

Pour passer d'une couche cachée à la suivante, l'ensemble des unités de traitement composant une couche calcule la somme pondérée par le coefficient du poids de leurs entrées et transmet généralement le résultat à une fonction d'activation non linéaire (voir Fig. 5.4). Il existe plusieurs fonctions d'activation ($tanh(z)$ ou $\frac{1}{1+exp^{-z}}$). ReLu (rectification linéaire : $f(z) = max(z, 0)$) reste l'une des fonctions les plus utilisées car elle permet notamment un apprentissage souvent plus

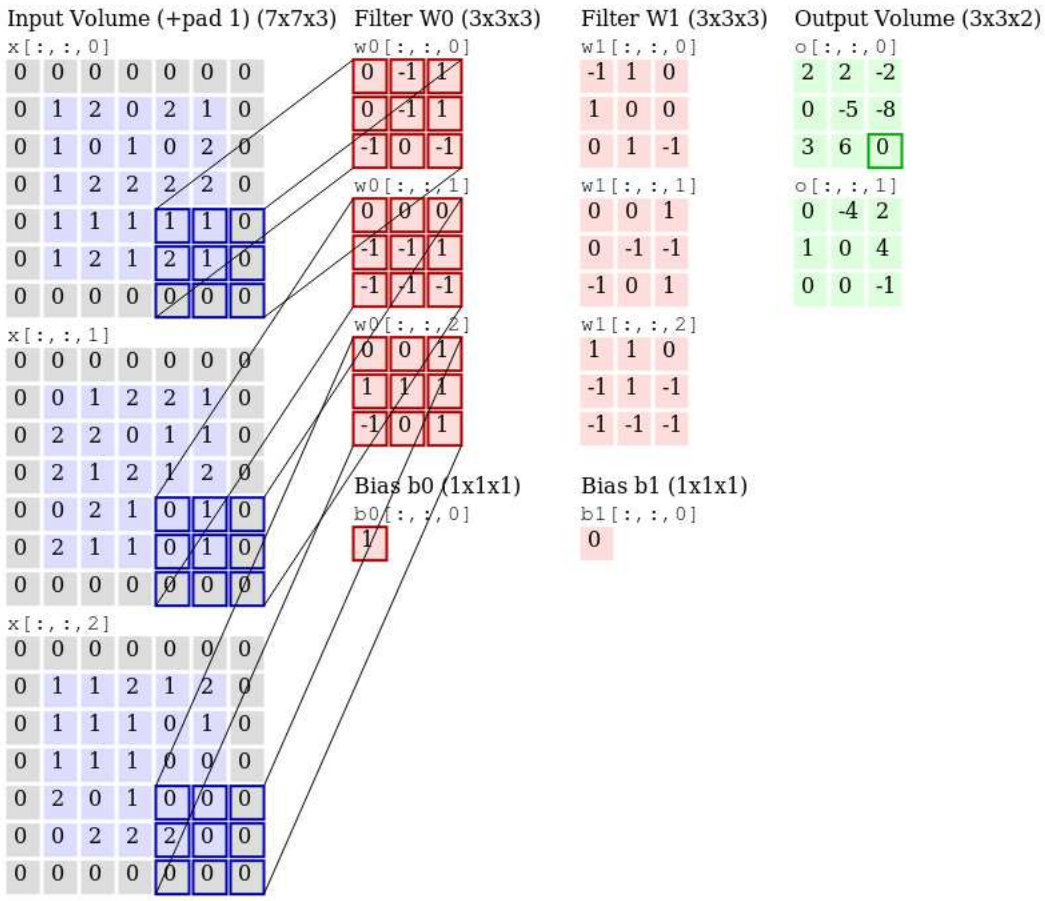


FIGURE 5.2 – Principe de convolution (source [5]).

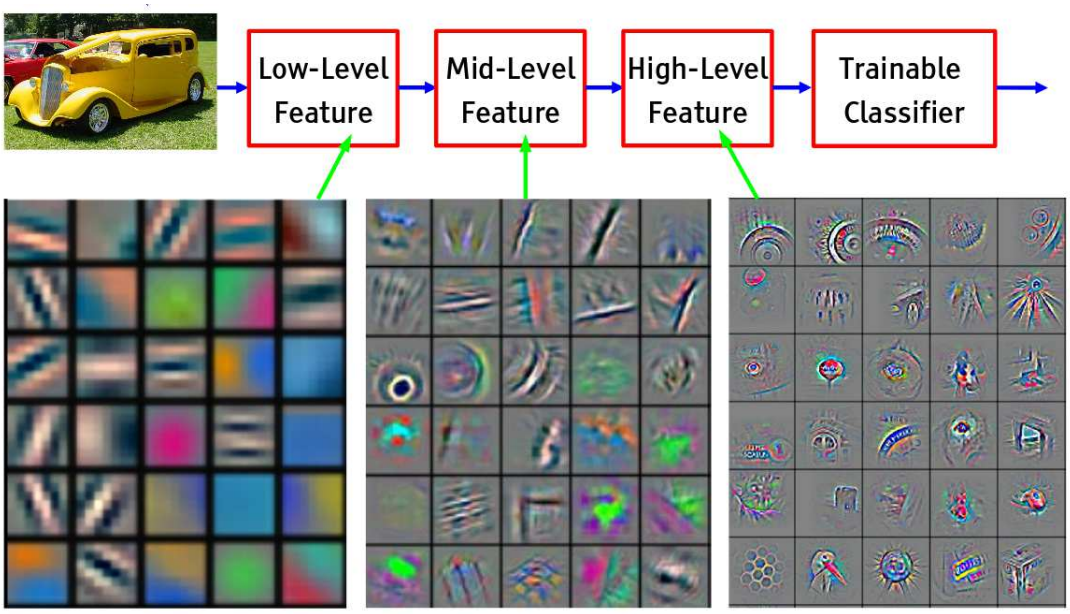


FIGURE 5.3 – Évolution des caractéristiques dans différentes couches d'un réseau de neurones (source [6]).

rapide.

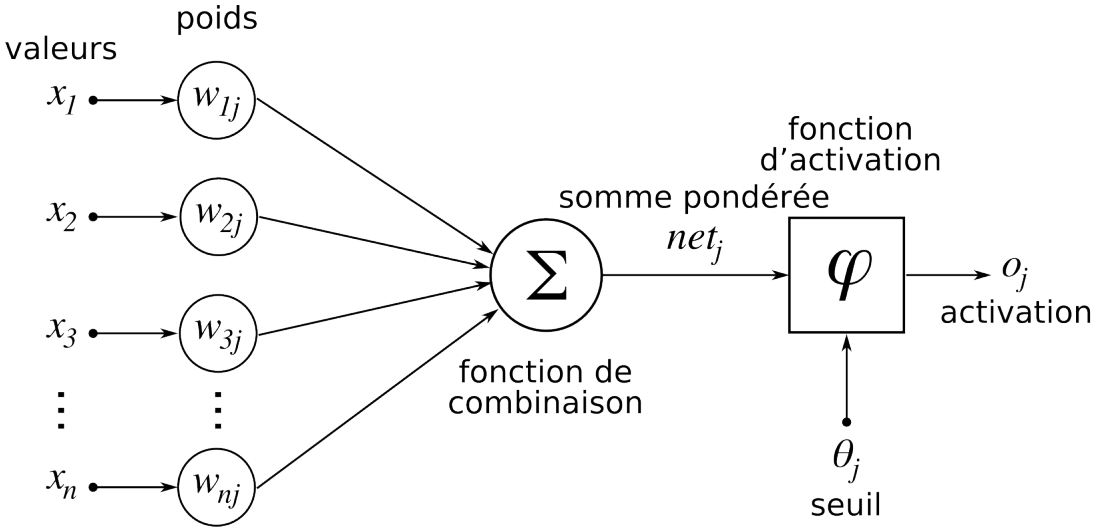


FIGURE 5.4 – Opérations effectuées par une unité de traitement (un neurone) dans une couche cachée d'un réseau de convolutions.

Après la fonction d'activation, on applique généralement une couche de sous-échantillonnage (Pooling) qui consiste à réduire la dimension des cartes de caractéristiques. Trois types de sous-échantillonnage sont souvent utilisés : maximum, minimum et moyenne. Il est à noter que l'on peut éventuellement définir son propre opérateur pour des tâches particulières. Cependant, plusieurs paramètres doivent être fixés (taille du filtre, recouvrement, etc.). La figure 5.5 montre un exemple de "max-pooling" avec un filtre de taille 2x2 et un pas de 2.

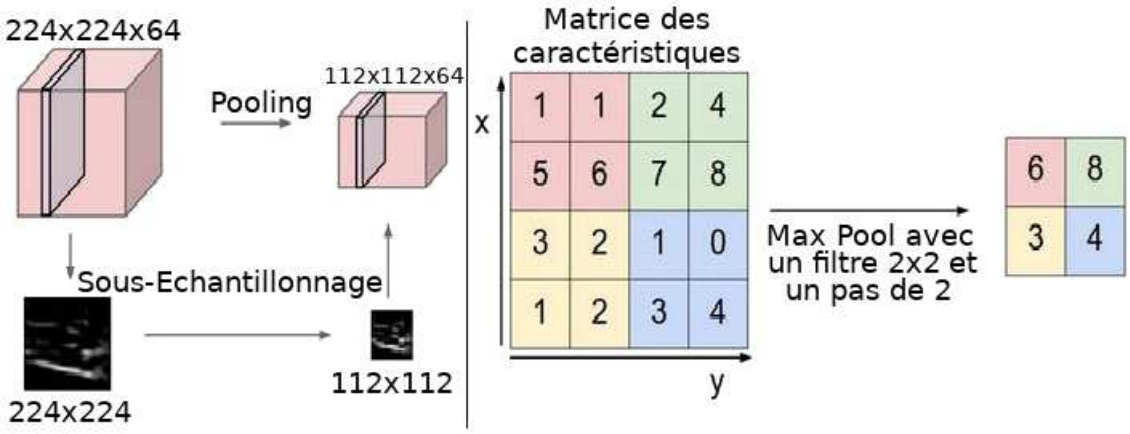


FIGURE 5.5 – Étape de sous-échantillonnage (Pooling) (source [5]).

Pour optimiser les paramètres du réseau CNN, un grand ensemble d'images étiquetées doit être disponible. Pendant la phase d'apprentissage, le réseau traite chaque image de la base d'apprentissage et produit une sortie sous la forme d'un vecteur de scores (une valeur pour chaque classe). La classe désirée doit obtenir le plus haut score (voir Fig. 5.6). Une fonction de perte ("loss function"), généralement de type descente de gradient stochastique (SGD), permet de mesurer l'erreur (ou la distance) entre les scores obtenus et les scores souhaités. A partir de cette fonction, le réseau modifie ses poids pour réduire les erreurs de prédiction.

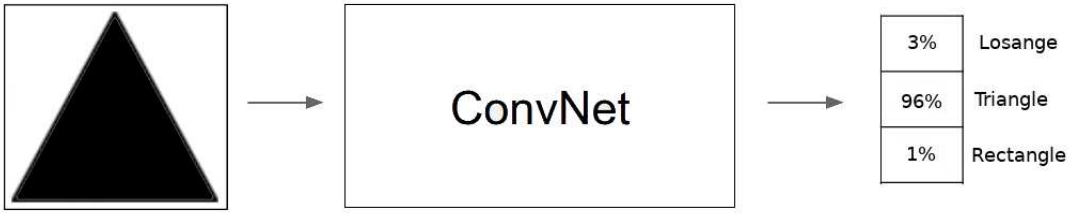


FIGURE 5.6 – Système classant des losanges, triangles et rectangles. Le plus haut score est obtenu pour la forme triangle. La prédiction du réseau est correcte.

Pour modifier de façon efficace les valeurs des poids du réseau, la méthode de rétro-propagation du gradient est souvent utilisée. Cette procédure permet de calculer le gradient d'une fonction de perte par rapport aux poids associés aux différentes couches en utilisant une dérivée en chaine. L'équation de rétro-propagation peut être appliquée à plusieurs reprises pour propager les gradients à travers toutes les couches, en commençant par la couche de sortie jusqu'à la couche d'entrée. Une fois les gradients obtenus, il est facile de calculer les gradients par rapport aux poids de chaque module.

La formule générale pour la mise à jour d'un poids dans un réseau de neurones s'écrit :

$$w_{i,j} \leftarrow w_{i,j} + \alpha a_i \Delta[j] \tag{5.1}$$

avec :

w : poids.

i : indice correspondant au neurone.

j : indice correspondant à la connexion.

α : coefficient d'apprentissage du réseau.

$a_i = \frac{\partial}{\partial w_{i,j}} in_j$: dérivée de la fonction d'activation avec in l'entrée de la fonction d'activation.

$\Delta[j] = -\frac{\partial}{\partial in_j} Loss(y_t, h_w(x_t))$: dérivée de la fonction de perte $Loss$ avec y_t la valeur souhaitée et $h_w(x_t)$ la valeur prédite par le neurone x à la position t.

Plusieurs couches de convolution, de fonctions d'activation et de "Pooling" (Conv->ReLU->Pool) sont généralement empilées, suivies de couches entièrement connectées (Fully Connected) (voir Fig 5.7). Les unités de traitement dans une couche entièrement connectée ont des connexions vers toutes les sorties de la couche précédente. La rétro-propagation des gradients est ensuite effectuée pour modifier la valeur des poids dans les différentes couches permettant d'améliorer la prédiction du réseau de neurones.

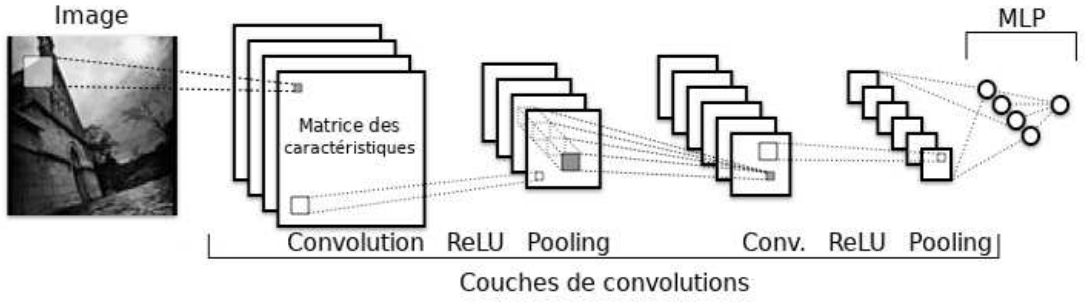


FIGURE 5.7 – Structure classique d'un réseau de neurones convolutionnel.

Après la phase d'apprentissage, la performance du système est mesurée sur un ensemble d'exemples appelé base de test. Cela permet d'évaluer la capacité du modèle à produire des réponses sensées sur de nouvelles données.

Les inconvénients majeurs de ce type de méthode sont la nécessité d'avoir une base de données conséquente et de disposer de capacités de calcul relativement élevées. Pour pallier à ces problèmes, un certain nombre de modèles déjà appris ont été repris et utilisés dans d'autres contextes. On parle de "Transfer Learning", notion qui est présentée dans la section suivante.

5.2 Transfer Learning

Le Transfer Learning est défini comme l'exploitation d'un réseau de neurones profond déjà entrainé sur une grande base de données, comme la base ImageNet [100]. La valeur des poids d'un modèle peut être utilisée soit comme paramètres d'initialisation pour entrainer ce même réseau sur une nouvelle base de données (Transfer Learning avec Fine-tuning) ou soit exploiter directement ce réseau sans appliquer une nouvelle étape d'apprentissage (Transfer Learning avec gel des poids). Le Fine-tuning est une méthode permettant d'ajuster la valeur des poids du modèle pré-entrainé via une nouvelle phase d'apprentissage. La seule modification nécessaire du réseau consiste à adapter la dernière couche. Cette dernière couche doit contenir le même nombre d'unités de traitement que de classes de la nouvelle base d'apprentissage.

Avec un réseau pré-entrainé, il est également possible d'extraire des descripteurs à la sortie d'une de ses couches pour alimenter un classifieur. Chaque image de la base de données est ainsi transformée en un vecteur de caractéristiques qui est utilisé pour entraîner un nouveau classifieur (voir Fig. 5.8). Cela permet d'obtenir des caractéristiques à partir d'une image, sans appliquer des méthodes de traitement d'images classiques.

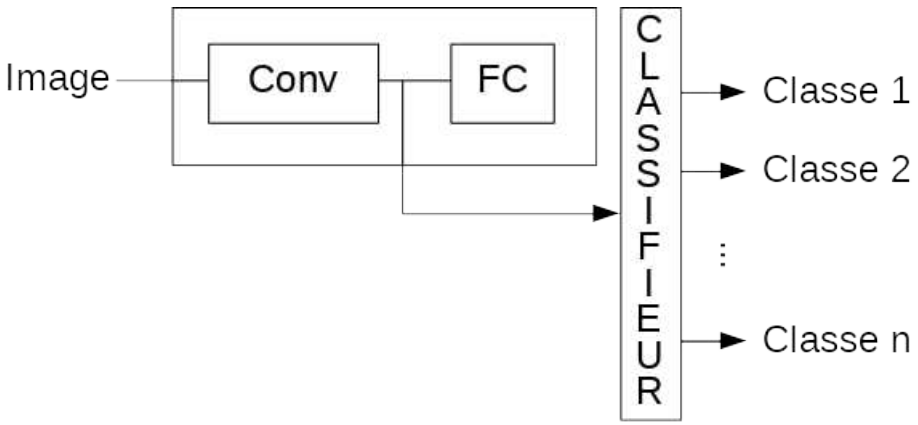


FIGURE 5.8 – Principe du Transfer Learning.

Pour utiliser un réseau déjà pré-entrainé, il est important de tenir compte de la taille de la nouvelle base de données et de sa similarité avec la base de données originale (ayant servie à l'apprentissage du réseau de neurones). Il existe quatre scénarios majeurs :

— La nouvelle base de données est de taille réduite et similaire à la base d'origine : Les caractéristiques des couches supérieures du réseau de neurones sont également pertinentes pour cette nouvelle base de données. Par conséquent, le réseau de neurones peut servir directement pour la classification.

— La nouvelle base de données est relativement grande et similaire à la base d'origine : Étant donné que nous disposons de plus de données, nous pouvons affiner le réglage des poids sans pour autant avoir un réseau trop spécialisé (Fine Tuning).

— La nouvelle base de données est de taille plus petite et dispose d'un contenu différent : Il est préférable d'utiliser un classifieur linéaire à partir des descripteurs obtenus par les premières couches du réseau de neurones.

— La nouvelle base de données est relativement grande et dispose d'un contenu différent : nous avons suffisamment de données pour affiner l'intégralité des poids du réseau de neurones (Fine Tuning).

Dans notre cas, le contenu de notre base est différent des images naturelles servant à l'apprentissage des réseaux standards. Nos images binaires représentent des motifs géométriques. Nous disposons d'une base de données suffisamment grande pour appliquer le Fine Tuning. Nous allons donc utiliser un réseau avec un faible nombre de couches.

Dans la section suivante, nous présentons trois des modèles pré-entrainés les plus utilisés dans la littérature : AlexNet, ResNet et VGG.

5.2.1 AlexNet

Alex Krizhevsky, Ilya Sutskever et Geoffrey Hinton sont les concepteurs du réseau AlexNet [7] (voir Fig. 5.9). Ce réseau a gagné la compétition Imagnet Large Scale Visual Recognition Challenge (ILSVRC) en 2012 (Défi de reconnaissance visuelle à grande échelle). C'est le premier réseau à atteindre un taux d'erreurs Top 5 de 15.4% (l'erreur Top 5 est le taux auquel, pour une image donnée, le vecteur de sortie du CNN n'a pas la classe de l'image parmis les 5 probabilités les plus élevées). Ce réseau est constitué de 5 couches de convolution et de 3 couches entièrement connectées. La fonction d'activation ReLU a été utilisée. Ce réseau utilise également le Dropout [8] qui permet de supprimer des connections entre les neurones des couches entièrement connectées, évitant le sur-apprentissage lors de l'entrainement du réseau.

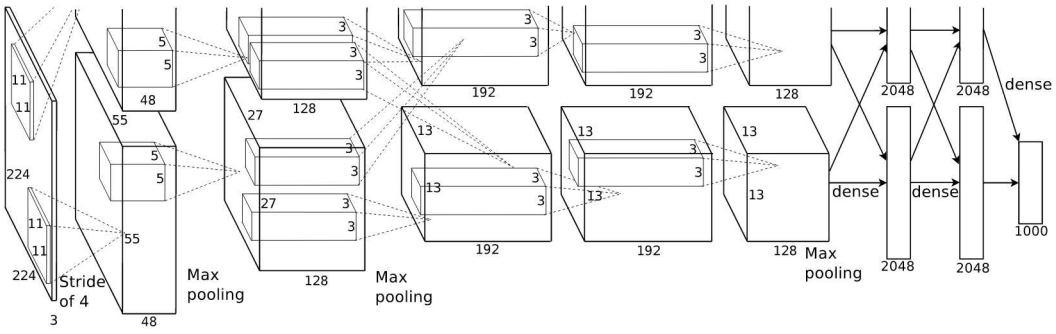
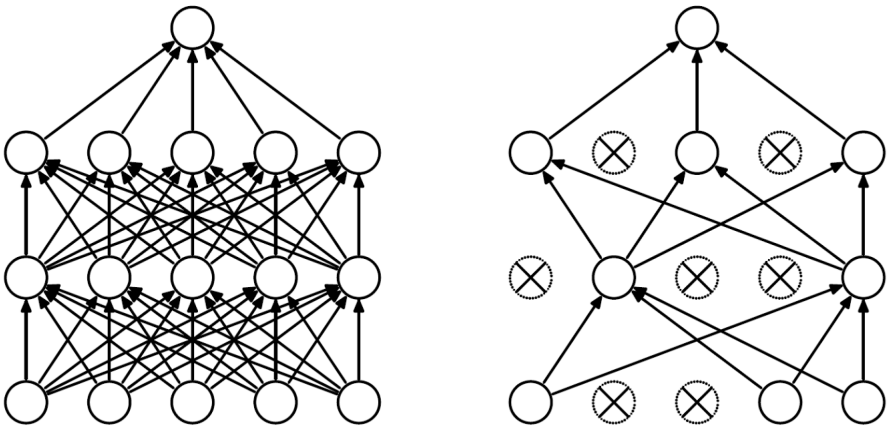


FIGURE 5.9 – Structure du réseau AlexNet (source [7]).

Les différentes techniques utilisées (ReLU, Dropout, etc.) sont toujours d'actualité et sont souvent utilisées. Ce réseau a montré les capacités des CNN pour la tâche de classification. Depuis, de nombreuses équipes de recherche travaillent sur ce type d'approche.

5.2.2 VGG

Ce modèle, créé en 2014 par Karen Simonyan et Andrew Zisserman de l'Université d'Oxford [9], a obtenu un taux d'erreur Top-5 de 7,3% (VGG 19 couches) lors de la compétition ILSVRC 2014. Il est composé de 19 couches qui utilise des filtres de convolution de 3x3 pixels. L'utilisation de filtres de taille 3x3 diffère des



(a) Standard Neural Net (b) After applying dropout.

FIGURE 5.10 – Dropout : suppression de connexions entre les neurones entièrement connectés (source[8]).

filtres 11x11 de la première couche de convolution puis du 5x5 de la deuxième couche du réseau AlexNet.

Si on combine deux couches de convolution 3x3, on obtient une unité de traitement effectif de 5x5. On peut donc obtenir la taille de filtre souhaité à partir d'un filtre 3x3 et diminuer ainsi le nombre de paramètres. En effet, si nous supposons que la profondeur des couches est de C canaux, alors on peut voir que la couche de convolution composée de filtres 7x7 contiendrait $C \times (7 \times 7 \times C) = 49C^2$ paramètres, tandis que les trois couches de convolutions composées de filtres 3x3 ne contiendraient que $3 \times (C \times (3 \times 3 \times C)) = 27C^2$ paramètres. De plus, avec trois couches de convolution, il est possible d'utiliser trois couches d'activation au lieu d'une afin d'obtenir de meilleurs descripteurs.

À mesure que la taille spatiale des volumes d'entrée de chaque couche diminue (résultat des couches de convolution et de sous-échantillonnage), la profondeur augmente avec le nombre de filtres lorsque l'on passe d'une couche à la suivante. Il est intéressant de remarquer que le nombre de filtres double après chaque couche de sous-échantillonnage. Cela renforce l'idée de réduire les dimensions spatiales tout en augmentant la profondeur des couches. Plusieurs configurations ont été

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
Input (224x224 RGB image)					
conv3-64	conv3-64 **LRN**	conv3-64 **conv3-64**	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
Maxpool					
conv3-128	conv3-128	conv3-128 **conv3-128**	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
Maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 **conv1-256**	conv3-256 conv3-256 **conv3-256**	conv3-256 conv3-256 conv3-256 **conv3-256**
Maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
Maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
Maxpool					
FC-4096					
FC-4096					
FC-1000					
Soft-Max					

FIGURE 5.11 – Configurations des différentes versions du réseau VGG. La profondeur des configurations augmente de gauche (A) à droite (E), à mesure que d'autres couches sont ajoutées (les couches ajoutées sont indiquées en gras). Les paramètres de la couche de convolution sont désignés par "conv.(taille du champ réceptif)-(nombre de canaux)". La fonction d'activation ReLU n'est pas montrée (source [9]).

proposées (voir Fig. 5.11).

5.2.3 ResNet

En 2015, Microsoft Research Asia a conçu le ResNet [10]. Ce réseau a une nouvelle architecture de 152 couches qui a permis d'établir de nouveaux records dans la classification, la détection et la localisation. Il a remporté la compétition ILSVRC 2015 avec un taux d'erreur Top-5 de 3,6% (le taux d'erreur Top-5 d'un humain est entre 5 et 10%). De telles performances viennent de l'ajout d'un "module résiduel" (voir Fig. 5.12) dont l'idée principale est que l'entrée x passe par une série de couches de convolution-ReLU-convolution que l'on note $F(x)$. Ce résultat est ensuite ajouté à l'entrée d'origine x ($H(x) = F(x) + x$). Dans les réseaux convolutifs traditionnels $H(x) = F(x)$. Fondamentalement, le module calcule une légère modification de l'entrée x pour obtenir une représentation légèrement modifiée. Dans les réseaux traditionnels, on ne conserve aucune information sur l'entrée x.

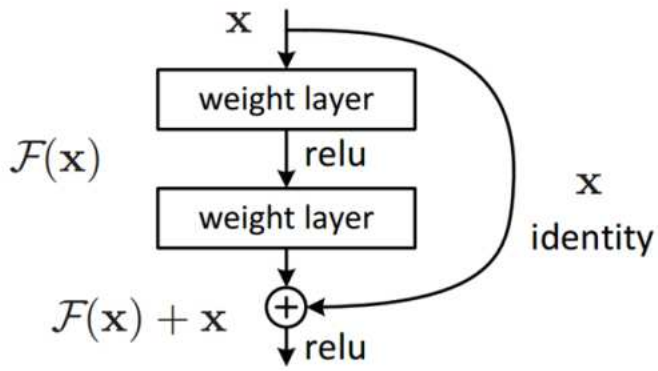


FIGURE 5.12 – Représentation d'un "module résiduel" (source [10]).

5.3 Résultats Expérimentaux

Afin de comparer les performances obtenues à celles du chapitre précédent, une série de tests a été réalisées. Les résultats présentés dans cette section sont basés sur le protocole suivant :

— Utilisation de la base 2 : 888 images binaires de décors de tessons répartis en 4 classes.

— Répartition de la base : 66% pour l'apprentissage et 33% pour la prédiction obtenues aléatoirement sans recouvrement.

— Taux moyen de bonne classification obtenu sur 20 jeux de données aléatoires.

— La bibliothèque utilisée est PyTorch [101].

5.3.1 Conception d'un réseau

Nous avons conçu un réseau en s'inspirant du modèle le plus simple. Ce réseau se compose de 4 couches de convolution et d'une couche entièrement connectée. La profondeur des couches de convolution est respectivement de 64, 192, 64, 160 filtres (voir Fig. 5.13). Ces valeurs de profondeurs ont été obtenues par des tests empiriques en faisant varier la profondeur de chaque couche. Les filtres de convolutions sont de taille 5x5 pixels. Chaque couche de convolution est suivie de la fonction d'activation ReLU, d'une normalisation et d'un sous-échantillonnage (max pooling : filtre de taille 2x2 pixels avec un pas de 2 pixels). L'apprentissage du réseau est fait sur 25 itérations.

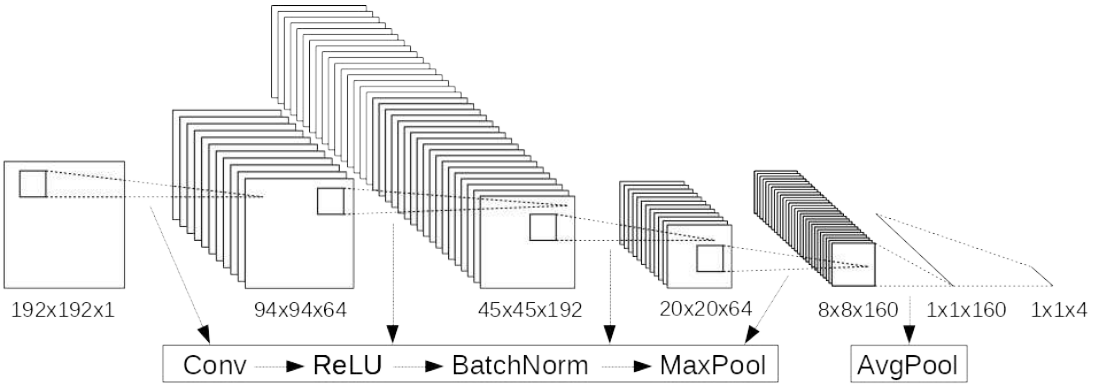


FIGURE 5.13 – Structure du réseau développé.

TABLE 5.1 – Taux moyen de bonne classification obtenu pour le modèle CNN proposé.

	Moyenne (%)
Modèle CNN proposé	71,97 ± 12,18

Avec ce réseau, nous avons obtenu un taux moyen de bonne classification de 71,97% et une variance des pourcentages de classification de 12.18%. Ce résultat

est inférieur au taux obtenu avec l'approche Blob-SIFT dans le chapitre précédent (84.14% / 2.93%). Comme énoncé précédemment, la taille de la base d'apprentissage joue un rôle important dans l'estimation des paramètres. Ayant uniquement 586 images pour l'apprentissage, les résultats ne sont pas probants. Afin d'améliorer les performances, nous proposons d'utiliser un des modèles de la littérature (Transfer Learning avec Fine Tuning).

5.3.2 Transfer Learning (Fine Tuning)

Nous avons sélectionné trois modèles (AlexNet, VGG11 et ResNet18) de la littérature en modifiant uniquement la structure des couches FC (pour passer de 1000 classes à 4). Pour chaque jeux de données, un apprentissage sur 25 itérations a été appliqué. Les résultats sont présentés dans la table 5.2 :

TABLE 5.2 – Taux moyen de bonne classification avec les réseaux ResNet18, VGG11 et AlexNet.

	Moyenne (%)
ResNet18	83.32 ± 4.80
VGG11	81.50 ± 6.40
AlexNet	77.24 ± 7.45

Une nette amélioration est observée avec l'utilisation d'un modèle pré-entrainé (e taux moyen de bonne classification passe de 71,97% à 83.32%). Parmi les 3 modèles, les meilleures performances ont été obtenues en utilisant le modèle ResNet18, aussi bien en termes de pourcentage de bonne classification qu'en termes de dispersion des pourcentages de classification obtenus. Cela est surement due à la spécificité du réseau ResNet18 (structure basée sur des modules résiduels).

Afin d'optimiser l'utilisation du réseau ResNet18, nous proposons de reprendre ce réseau et de modifier la partie FC. Le premier test a consisté à ajouter une couche cachée afin de réduire le passage de 512 neurones (vecteur de sortie de la dernière couche de convolution) à 4 neurones (nombre de classes). Cela permet d'avoir une couche intermédiaire atténuant cette transition. La couche cachée ajou-

tée est constituée de 170 neurones. Nous nommerons ce réseau ResNet18-FC170. Les résultats obtenus sont présentés par la table 5.3 :

TABLE 5.3 – Taux moyen de bonne classification obtenu avec le réseau ResNet18-FC170.

	Moyenne (%)
ResNet18-FC170	84.53 ± 4.50
ResNet18	83.32 ± 4.80

Comme on peut le constater, l'ajout de cette couche supplémentaire a permit une meilleure séparation des classes avec un taux moyen de bonne classification qui a augmenté de 1% tout en diminuant la variance des pourcentages de bonne classification.

Le deuxième test a consisté à remplacer directement le classifieur du CNN par un autre classifieur (SVM et arbre de décision). Pour le classifieur SVM, nous avons utilisé un noyau de type χ^2. Pour le classifieur de type arbre de décision, nous avons utilisé les gradients boosting (Librairie XGBoost [83]). Les résultats sont donnés dans la table 5.4 :

TABLE 5.4 – Comparaison des classifieurs.

	Moyenne (%)
SVM	87.94 ± 2.20
XGBoost	87.29 ± 1.60
ResNet18-FC170	84.53 ± 4.50

Le changement de classifieur a permit d'améliorer les performances d'un peu plus de 3% dans le meilleur des cas avec une variance des pourcentages de classification qui a diminuée. Afin de mieux voir les erreurs entre les classses, la matrice de confusion obtenues pour le classifieur de type SVM est présentée dans la table

5.5 (Rappel, classe A : losanges, classe C-G : bâtons ou carrés sur deux registres, classe H : carrés sur trois registres, classe L : chevrons).

TABLE 5.5 – Matrice de confusion pour le classifieur de type SVM en sortie du CNN.

		Classes Prédites			
Classes Réelles	**%**	**A**	**C-G**	**H**	**L**
	A	**95.50**	0.82	1.00	2.68
	C-G	2.01	**79.81**	15.58	2.60
	H	3.33	9.88	**86.33**	0.46
	L	3.97	1.09	0.20	**94.74**

Les meilleures performances ont été obtenues pour la classe A et L avec des taux moyen de bonne classification supérieurs à 94%. Les confusions les plus importantes ont été obtenues entre les classes C-G et H. Cela s'explique notamment par la proximité de ces classes. En effet, certains des échantillons de ces classes respectives ne diffèrent que par le nombre de registres. On constate aussi que la tendance (des confusions) est relativement similaire à celle de l'approche Blob-SIFT+BoW.

5.4 Approche combinée

Les résultats obtenus avec les descripteurs issus de l'approche Blob-SIFT et les descripteurs du ResNet18-FC170 sont relativement proches. Le descripteur Blob-SIFT donne des caractéristiques "locales", tandis que pour le ResNet18, l'entrée est l'image complète du décor, on a donc un descripteur "global". Afin d'améliorer les performances, nous proposons de combiner ces deux approches pour obtenir un vecteur de caractéristiques de taille 340 (voir Fig. 5.14). Ce vecteur est utilisé comme entrée de différents classifieurs (SVM avec un noyau χ^2 et arbre de décision). Les résultats sont présentés dans la table 5.6 :

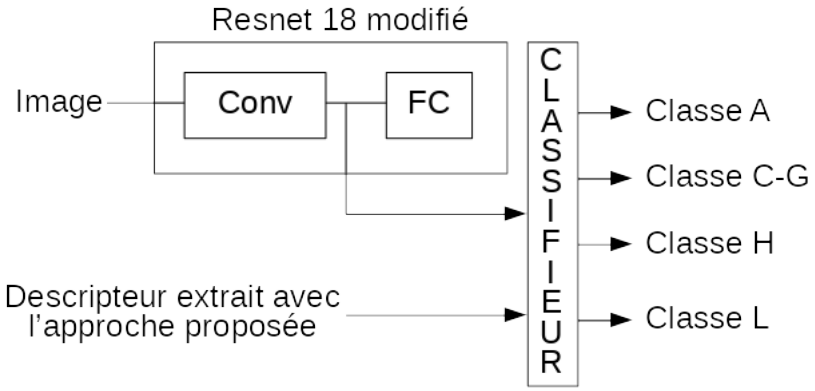


FIGURE 5.14 – Fusion des descripteurs pour la classification.

TABLE 5.6 – Taux moyen de bonne classification avec la combinaison des descripteurs de issus de l'approche Blob-SIFT et des descripteurs issus du ResNet18-FC170.

	Moyenne (%)
SVM	89.22 ± 2.40
XGBoost	88.17 ± 3.56

En combinant les informations locales et globales, la classification est nettement améliorée avec un taux moyen de bonne classification proche de 90% et une variance des pourcentages de bonne classification qui est faible. La matrice de confusion pour la classification de type SVM est donnée dans la table 5.7.

Les confusions ont diminué avec les mêmes remarques que pour la matrice de confusion présentée précédemment (les confusions les plus importantes sont toujours entre les classes C-G et H).

A travers ce chapitre nous avons comparé une approche basée sur une extraction de caractéristiques à une approche basée sur un apprentissage profond. L'analyse des résultats nous a permit de proposer une méthode hybride combinant les vecteurs de caractéristiques obtenus par les deux approches. Les performances ont été ainsi améliorées.

TABLE 5.7 – Matrice de confusion de la classification par SVM de la concaténation des vecteurs de caractéristiques issus des descripteurs Blob-SIFT et des descripteurs issu du ResNet18-FC170.

		Classes Prédites			
Classes Réelles	**%**	**A**	**C-G**	**H**	**L**
	A	**96.85**	0.20	0.80	2.15
	C-G	2.87	**82.52**	12.45	2.16
	H	2.10	9.63	**88.22**	0.05
	L	5.26	2.18	0.40	**92.16**

Conclusion

Dans le cadre du projet ARCADIA, nos travaux de recherche se sont intéressés à l'analyse automatique de décors gravés à la molette sur les fragments de céramiques trouvés sur le site de la Médecinerie dans le Loiret. Pour ce faire, nous avons développé une chaîne complète de traitements depuis la numérisation de ce patrimoine archéologique jusqu'à la classification automatique des décors selon leur style (carré, losange, chevrons, oves, etc).

Une première base constituée de 377 scans 3D de tessons répartis en 5 classes expertisées par l'archéologue a permis de mettre au point les prétraitements, puis paramétrer et comparer différents descripteurs globaux et locaux sur nos décors par le biais de deux classifieurs : séparateurs à vaste marge (SVM) et arbre de décision (XGBoost). La base a ensuite été élargie à 888 tessons répartis en 4 classes et utilisée pour tester une approche par apprentissage profond.

A partir du nuage de point 3D, une carte de profondeurs est obtenue. Le passage de la 3D à la 2D facilite les traitements. Pour diminuer l'effet de la courbure et faire ressortir le décor, nous avons calculé la carte des variances locales. A partir de ces images, nous avons utilisé le détecteur FAST et appliqué la méthode de clustering DBSCAN pour repérer les zones de textures les plus denses présentes sur l'image (portions du décor). Les clusters ou zones denses de points d'intérêt localisent la région saillante où se situe le décor sur le tesson et permettent d'optimiser la phase de binarisation. On peut alors extraire de la carte de profondeurs une empreinte binaire du décor gravé semblable à l'encrage manuel réalisé par l'archéologue.

Sur cette base de décors binarisés, une étude comparative a confronté plusieurs descripteurs globaux et locaux de l'état de l'art avec un classifieur SVM

ou XGBoost. Ces deux classifieurs présentent des performances similaires. La caractérisation globale des décors par les filtres de Gabor plafonne à 61.04%. Les approches classiques en indexation d'images, associant l'extraction de descripteurs denses sur des grilles régulières et occurrence de sac-de-mots visuels, se montrent plus efficaces avec 81.22% obtenus par le descripteur SIFT normalisé.

La nouvelle caractérisation des décors proposée, Blob-SIFT, surpasse les autres méthodes récentes parmi les plus performantes de l'état de l'art pour générer des dictionnaires visuels et permet d'obtenir un score de classification de 84.76%. Le choix d'extraire les descripteurs SIFT sur une grille irrégulière adaptée à chaque décor par des points d'intérêt localisés au centre de chaque éléments du décor s'avère donc être la meilleure stratégie. Cette approche permet à la fois de réduire considérablement les ressources mémoires nécessaires par rapport à une grille dense et d'améliorer les performances de classification. Enfin, les tests réalisés démontrent que l'empreinte binaire est plus efficace pour caractériser les décors que la carte de profondeurs ou la carte des variances locales. L'étape de focalisation sur la zone saillante est très importante puisque en découleront le résultat de la binarisation de la carte de profondeurs et donc directement les performances de la classification.

L'approche de classification supervisée par extraction de caractéristiques a été comparée à une approche d'apprentissage profond sur la seconde base. Cette base restant insuffisante pour entrainer correctement un réseau nouvellement créé, nous avons opté pour le Transfer Learning. Trois réseaux ont été testés avec un nombre de couches faible car nos images, composées de formes binaires géométriques, ne nécessitent pas plus de complexité. La meilleure configuration avec ResNet18 a donné un score de 83.32%. La modification de ce réseau ajoutant une couche entièrement connectée a permis d'améliorer les résultats en obtenant 84.53% de bonnes classifications, un score proche de celui obtenu avec une classification SVM avec le descripteur Blob-SIFT.

Enfin, nous avons fusionné le vecteur de caractéristiques locales extrait par le descripteur Blob-SIFT et la caractérisation globale du décor issus de l'apprentissage profond. Cette approche hybride a permis une nette amélioration du taux de

classification moyen avec un score de 89.22%. Les principales confusions restantes concernent les classes carrées sur deux et trois registres.

Les travaux futurs porteront d'abord sur l'agrandissement de la base de données des tessons. Au cours de ces travaux, l'accès au bâtiment où sont stockés les tessons a été malheureusement fermé pour des raisons techniques. Dans l'attente de la réouverture des locaux, nous pouvons néanmoins augmenter la base de données artificiellement en appliquant des transformations géométriques sur tout ou partie de nos images.

L'étape d'extraction de la région saillante nous a permis d'obtenir un taux moyen de bonne classification probant. Cependant, les tests préliminaires de classification réalisés en utilisant directement la vérité terrain (extraction manuelle de la région saillante) montrent un gain de bonne classification d'un peu plus de 2%. Ainsi, cette étape étant cruciale, il serait intéressant de l'approfondir à travers l'exploitation notamment de méthodes d'extraction des régions saillantes dont certaines sont basées sur des réseaux convolutifs.

Pour évaluer les performances des modèles CNN utilisés, plusieurs modalités ont été testées. Les meilleures performances ont été obtenues en utilisant l'image du décor binaire. Comme nous l'avons constaté dans le chapitre 4, les erreurs de classification obtenues sont souvent dues à une binarisation imparfaite des tessons. Plusieurs pistes d'amélioration sont possible. Nous envisageons d'utiliser des méthodes plus efficaces basées notamment sur des réseaux profonds qui permettront à la fois d'extraire la région saillante et de binariser l'image.

Pour l'étape de classification, une version multi-noyaux du SVM, communément appelée Multiple Kernel Learning (MKL) pourra être testée. L'avantage de cette approche est de pouvoir combiner plusieurs noyaux. Ces derniers peuvent être de même types avec la possibilité d'avoir différents paramètres pour chaque noyau ou de type différent. La combinaison de plusieurs noyaux pourra permettre une meilleure séparation des données.

Publications liées à la thèse

5.5 Journal

Debroutelle T., Chetouani A., Treuillet S., Martin L., Exbrayat M., and Jesset S. Automatic classification of ceramic sherds with relief motifs. *Journal of Electronic Imaging*, 26(2 :023010, 03 2017.

5.6 Conférences Internationales

Debroutelle T., Chetouani A., Treuillet S., Martin L., Exbrayat M., and Jesset S. Classification of friezes engraved on ceramic sherds from 3d scans. *IEEE International Symposium on Signal Processing and Information Technology*, pages 218–222, 03 2016.

Debroutelle T., Chetouani A., Treuillet S., Martin L., Exbrayat M., and Jesset S. Automatic pattern recognition on archaeological ceramic by 2d and 3d image analysis : A feasibility study. *Image Processing Theory, Tools and Applications*, pages 224–228, 11 2015.

5.7 Conférences Nationales

Martin L., Exbrayat M., Debroutelle T., Chetouani A., Treuillet S., and Jesset S. Recherche de groupes parallèles en classification non-supervisée. *Extraction et Gestion des Connaissances*, pages 69–80, 2016.

Debroutelle T., Chetouani A., Treuillet S., Martin L., Exbrayat M., and Jesset S. Extraction et classification de motifs de tessons de céramique. *Reconnaissance de*

Formes et l'Intelligence Artificielle (RFIA), 2016.

Debroutelle T., Chetouani A., Treuillet S., Martin L., Exbrayat M., and Jesset S. Détection automatique de motifs céramiques archéologiques par analyse d'images 2d. *GRETSI*, 2015.

Janvier R., Debroutelle T., Chetouani A., Treuillet S., Exbrayat M., Martin L., and Jesset S. Etude de faisabilité d'une reconnaissance automatique de motifs céramiques archéologiques par analyse d'images 2d et 3d. *Congrès des jeunes chercheurs en vision par ordinateur - ORASIS*, 06 2015.

Bibliographie

[1] Edward Rosten and Tom Drummond. Machine learning for high-speed corner detection. *In European Conference on Computer Vision*, 1 :430–443, 2006.

[2] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 2 :91–110, 2004.

[3] Li Fei-Fei, Rob Fergus, and Antonio Torralba. Short courses on recognizing and learning object categories. *International Conference on Computer Vision (ICCV)*, 2005.

[4] Anna Bosch, Andrew Zisserman, and Xavier Munoz. Scene classification via plsa. *European Conference on Computer Vision (ECCV)*, 2006.

[5] Cs231n convolutional neural networks for visual recognition. `http://cs231n.github.io/convolutional-networks/`.

[6] M.D. Zeiler and R Fergus. Visualizing and understanding convolutional networks. *European Conference on Computer Vision(ECCV)*, 8689 :818–833, 01 2013.

[7] Alex Krizhevsky. One weird trick for parallelizing convolutional neural networks. *CoRR*, abs/1404.5997, 2014.

[8] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout : A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15 :1929–1958, 06 2014.

[9] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *CoRR*, abs/1409.1556, 2014.

[10] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *CoRR*, pages 770–778, 06 2016.

[11] Jesset S. La diffusion dans l'orléanais des productions de l'atelier d'orléans-saran du 6e au 9e siècle. *Mémoire de DEA. Université de Tours*, 1995.

[12] Cribellier C., Jesset S., and Couvin F. Aperçu des décors sur céramiques en région centre de la tène à la fin de la période carolingienne : éléments pour une synthèse diachronique. *Spécificités et diffusion de la céramique gallo-romaine en région Centre ; actualité des recherches céramiques*, pages 337–376, 2005.

[13] Line Pastor. Molettes et roulettes de potiers gallo-romains dans l'est de la gaule. *Revue archéologique de l'Est*, 55 :287–297, 01 2007.

[14] Jesset S. Fouille programmée de 2009 : Saran, lac de la médecinerie. *Revue archéologique du Loiret et de l'axe ligérien*, 33 :103–107, 2009.

[15] C. Maiza. Classification d'objets de révolutions : application aux poteries sigillées. *Thèse, Université Paul Sabatier - Toulouse 3*, 2008.

[16] H. Mara. Documentation of rotationally symmetric archaeological finds by 3d shape estimation. *Rapport technique*, 2006.

[17] A. Willis, D. Cooper, and X. Orriols. Accuately estimating sherd 3d surface geometry with application to pot reconstruction. *Conference on Computer Vision and Pattern Recognition Workshop*, 2003.

[18] K. Son, E. Almeida, and D. Cooper. Axially symmetric 3d pots configuration system using axis of symmetry and break curve. *Conference on Computer Vision and Pattern Recognition*, pages 257–264, 2013.

[19] S.Y. Zheng, R.Y. Huang, J. Li, and Z. Wang. Reassembling 3d thin fragments of unknown geometry in cultural heritage. *SPRS Anals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 5(2) :393–399, 2014.

[20] A. Ippolito, L. Senatore, G. Ceroli, and Belelli M. B. From survey to the representation of the model, a documentation of typological and chronological sequences of archaeological artefacts : Traditional and innovative approach. *Computer Application and Quantitative Methods in Archaeology*, 2014.

[21] M. Kampel and R. Sablatnig. Color classification of archaeological fragments. *International Conference on Pattern Recognition*, pages 771–774, 2000.

[22] F. Stanco, A. M. Gueli, D. Tanasi, and G. Stella. Computer graphics solutions for pottery colors specification. *European Conference on Colour in Graphics, Imaging, and Vision*, pages 97–101, 2013.

[23] M. Farjas, J. G. Rejas, T. Mostaza, and J. Zancajo. Deepening in the 3d modelling : Multisource analysis of a polychrome ceramic vessel through the integration of thermal and hyperspectral information. *International Conference on Computer Applications and Quantitative Methods in Archaeology*, pages 116–124, 2012.

[24] Q. Li-Ying and W. Ke-Gang. Kernel fuzzy clustering based classification of ancient-ceramic fragments. *International Conference on Information Management and Engineering*, pages 348–350, 2010.

[25] M. Abadi, M. Khoudeir, and S. Marchand. Gabor filter-based texture features to archaeological ceramic materials characterization. *Image and Signal Processing, Lecture Notes in Computer Science, 7340*, pages 333–342, 2012.

[26] D. Gabor. Theory of communication. *Journal of the Institution of Electrical Engineers*, 93 :429–459, 1946.

[27] P. Smith, D. Bespalov, A. Shokoufandeh, and P. Jeppson. Classification of archaeological ceramic fragments using texture and color descriptors. *Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pages 49–54, 2010.

[28] M. Makridis and P. Daras. Automatic classification of archaeological pottery sherds. *Journal on Computing and Cultural Heritage*, 5(4) :1–21, 2012.

[29] G. C. Guarnera, F. Stanco, D. Tanasi, and G. Gallo. Classification of decorative patterns in kamares pottery. *Spring Conference on Computer Graphics*, pages 20–23, 2011.

[30] C. Xu, J. Liu, and X. TANG. 2d shape matching by contour flexibility. *Pattern Analysis and Machine Intelligence*, 31(1) :180–186, 2009.

[31] Jun Zhou, Haozhou Yu, Karen Smith, Colin Wilder, Hongkai Yu, and Song Wang. Identifying designs from incomplete, fragmented cultural heritage objects by curve-pattern matching. *Journal of Electronic Imaging*, 26, 2017.

[32] R. Seulin, C. Stolz, D. Fofi, G. Millon, and F. Nicolier. Three-dimensional tools for analysis and conservation of ancient wooden stamps. *The Imaging Science Journal*, 54 :111–121, 2006.

[33] W Niblack. An introduction to digital image processing. *Prentice Hall, W. Niblack, Englewood Cliffs*, 01 1986.

[34] S. Marchand. Ibisa 3d : Image-base identification/search for archaeology using a three-dimensional coin model. *Computer Applications and Quantitative Methods in Archaeology*, pages 11–21, 2013.

[35] H. Anwar, S. Zambanini, and M. Kampel. A bag of visual words approach for symbols-based coarse-grained ancient coin classification. *The 37th Annual Workshop of the Austrian Association for Pattern Recognition*, 2013.

[36] S. Zambanini and M. Kampel. A local image descriptor robust to illumination changes. *Image Analysis, Lecture Notes in Computer Science*, pages 11–21, 2013.

[37] C. Liu, J. Yuen, and A. Torralba. Sift flow : Dense correspondence across scenes and its applications. *Pattern Analysis and Machine Intelligence*, 33(5) :978–994, 2011.

[38] S. Zambanini and M. Kampel. Classifying ancient coins by local feature matching and pairwise geometric consistency evaluation. *International Conference on Pattern Recognition*, 2014.

[39] S. Marchand, P. Desbarats, A. Vialard, F. Bechtel, A. Ben Amara, B. Cicuttini, J. P. Bost, A. Bresson, K. Kounk, and A. Beurive. Ibisa : Image-base identification/search for archaeology. *Virtual Reality, Archaeology and Cultural Heritage*, pages 57–60, 2009.

[40] Next engine. `http://www.nextengine.com`.

[41] Farjas M., Rejas Juan G., Mostaza T., and Zancajo J. Deepening in the 3d modelling : multisource analysis of a polychrome ceramic vessel through the integration of thermal and hyperspectral information. *Conference on Computer Applications and Quantitative Methods in Archaeology*, pages 116–124, 04 2011.

[42] Tucci G., Cini D., and Nobile A. Effective 3d digitization of archaeological artifacts for interactive virtual museum. *International Archives of the*

Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives, 38 :413–420, 09 2012.

[43] Polo M.E. and Felicisimo A. Analysis of uncertainty and repeatability of a low-cost 3d laser scanner. *Sensors (Basel, Switzerland)*, 12 :9046–54, 12 2012.

[44] Debroutelle T., Janvier R., Chetouani A., Treuillet S., Exbrayat M., Martin L., and Jesset S. Automatic pattern recognition on archaeological ceramic by 2d and 3d image analysis : A feasibility study. *Image Processing Theory, Tools and Applications*, pages 224–228, 11 2015.

[45] Janvier R., Debroutelle T., Chetouani A., Treuillet S., Exbrayat M., Martin L., and Jesset S. Etude de faisabilité d'une reconnaissance automatique de motifs céramiques archéologiques par analyse d'images 2d et 3d. *Congrès des jeunes chercheurs en vision par ordinateur - ORASIS*, 06 2015.

[46] Radu Bogdan Rusu and Steve Cousins. 3d is here : Point cloud library (pcl). *IEEE International Conference on Robotics and Automation (ICRA)*, 05 2011.

[47] Alexandru Telea. An image inpainting technique based on the fast marching method. *Journal of Graphics Tools*, 9 :23–34, 01 2004.

[48] JA Sethian. A marching level set method for monotonically advancing fronts. *Proceedings of the National Academy of Sciences of the United States of America*, 93 :1591–1595, 03 1996.

[49] John Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6 :679–698, 1986.

[50] Chris Harris and Mike Stephens. A combined corner and edge detector. *In Proc. of Fourth Alvey Vision Conference*, pages 147–151, 1988.

[51] H. Bay, T. Tuytelaars, and L. Van Gool. Surf : Speeded up robust features. *Computer Vision and Image Understanding (CVIU)*, 110(3) :346–359, 2008.

[52] Ethan Rublee, Vincent Rabaud, Kurt Konolige, and Gary R. Bradski. Orb : An efficient alternative to sift or surf. *International Conference on Computer Vision*, pages 2564–2571, 2011.

[53] Stefan Leutenegger, Margarita Chli, and Roland Siegwart. Brisk : Binary robust invariant scalable keypoints. *International Conference on Computer Vision*, pages 2548–2555, 2011.

[54] David Nistér and Henrik Stewénius. Linear time maximally stable extremal regions. *European Conference on Computer Vision*, pages 183–196, 2008.

[55] Martin Ester, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. *AAAI Press*, pages 226–231, 1996.

[56] Delaunay B. A density-based algorithm for discovering clusters in large spatial databases with noise. *Bulletin de l'Académie des Sciences de l'URSS, Classe des Sciences Mathématiques et Naturelles*, 6 :793–800, 1934.

[57] Martin L., Exbrayat M., Debroutelle T., Chetouani A., Treuillet S., and Jesset S. Recherche de groupes parallèles en classification non-supervisée. *Extraction et Gestion des Connaissances*, pages 69–80, 2016.

[58] MacQueen J. B. Some methods for classification and analysis of multivariate observations. *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability. 1. University of California Press.*, pages 281–297, 1967.

[59] J. Sauvola and M. Pietikäinen. Adaptive document image binarization. *Pattern Recognition*, 33 :225–236, 2000.

[60] Xigyu Pan. Caractérisation de motifs céramiques par analyse d'images. *Rapport de stage Polytech Orléans*, 2012.

[61] T Ojala, Matti Pietikäinen, and D Harwood. Performance evaluation of texture measures with classification based on kullback discrimination of distributions. *Proceedings of the 12th IAPR International Conference on Pattern Recognition*, 1 :582 – 585, 11 1994.

[62] S.X. Liao and M. Pawlak. On image-analysis by moments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(3) :254–266, 1996.

[63] M. Exbrayat and L. Martin. Explorer 3d : Manuel d'utilisation. 2013.

[64] Anna Bosch, Andrew Zisserman, and Xavier Munoz. Image classification using random forests and ferns. *International Conference on Computer Vision (ICCV)*, pages 1–8, 01 2007.

[65] Mian Zhou and Hong Wei. Face verification using gabor wavelets and adaboost. *In : Pattern Recognition, International Conference on 1*, pages 404–407, 2006.

[66] Krystian Mikolajczyk and Cordelia Schmid. A performance evaluation of local descriptors. *Transactions on PAMI*, 10 :91–110, 2004.

[67] Michael Calonder, Vincent Lepetit, and Pascal Fua. Brief : Binary robust independent elementary features. *European Conference on Computer Vision*, 12 2011.

[68] Alexandre Alahi, Raphael Ortiz, and Pierre Vandergheynst. Freak : Fast retina keypoint. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 510–517, 06 2012.

[69] Antti Hietanen, Jukka Lankinen, Joni-Kristian Kämäräinen, Anders Glent Buch, and Norbert Krüger. A comparison of feature detectors and descriptors for object class matching. *Neurocomputing*, 184 :3–12, 12 2016.

[70] Fei-Fei Li, Pietro Perona, and California Institute of Technology. A bayesian hierarchical model for learning natural scene categories. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2 :524–531, 2005.

[71] S. Josef and A. Zisserman. Efficient visual search of videos cast as text retrieval. *Pattern Analysis and Machine Intelligence (PAMI)*, 31(4) :591–605, 2009.

[72] A. P. Dempster, Laird N. M., and Rubin D. B. Maximum likelihood from incomplete data via the em algorithm. *Journal Of The Royal Statistical Society*, 39(1) :1–38, 1977.

[73] Jorge Sanchez, Florent Perronnin, Thomas Mensink, and Jakob Verbeek. Image classification with the fisher vector : Theory and practice. *International Journal of Computer Vision*, 105(3) :222–245, 12 2013.

[74] Herve Jegou, Florent Perronnin, Matthijs Douze, Jorge Sanchez, Patrick Perez, and Cordelia Schmid. Aggregating local image descriptors into compact codes. *IEEE transactions on pattern analysis and machine intelligence*, 34 :1704–1716, 12 2011.

[75] Olivier Chapelle, Patrick Haffner, and Vladimir N. Vapnik. Support vector machines for histogram-based image classification. *IEEE transactions on neural networks, a publication of the IEEE Neural Networks Council*, 10 :1055–1064, 09 1999.

[76] B. Scholkopf and A. J. Smola. Learning with kernels. *MIT Press*, 2002.

[77] Chih-Chung Chang and Chih-Jen Lin. Libsvm : a library for support vector machine. *ACM Transactions on Intelligent Systems and Technology*, 2 :1–27, 2011.

[78] L. Breiman, J. Friedman, Olshen R., and Stone C. Classification and regression trees. *Wadsworth and Brooks*, 1984.

[79] Manuel Fernandez-Delgado, Eva Cernadas, Senen Barro, and Dinani Amorim. Do we need hundreds of classiers to solve real world classication problems ? *Journal of Machine Learning Research*, 15 :3133–3181, 2014.

[80] Breiman L. Bagging predictors. *Machine Learning*, 26(2) :123–140, 1996.

[81] Breiman L. Random forests. *Machine Learning*, 45 :5–32, 2001.

[82] Y. Freund and R. E. Schapire. Experiments with a new boosting algorithm. *Machine Learning : Proceedings of the Thirteenth International Conference*, 45 :148–156, 1996.

[83] Tianqi Chen and Carlos Guestrin. Xgboost : A scalable tree boosting system. *The 22nd ACM SIGKDD International Conference*, pages 785–794, 08 2016.

[84] Andrea Vedaldi and Andrew Zisserman. Efficient additive kernels via explicit feature maps. *IEEE transactions on pattern analysis and machine intelligence*, 34 :480–92, 07 2011.

[85] Mohamed Fezari, Abadi Wassila, and Rachid Hamdi. Bag of visual words and chi-squared kernel support vector machine : A way to improve hand gesture recognition. *Proceedings of the International Conference on Intelligent Information Processing, Security and Advanced Communication*, 91 :1–5, 2015.

[86] A. Vedaldi and B. Fulkerson. Vlfeat : An open and portable library of computer vision algorithms. 2008.

[87] G. Bradski. The opencv library. *Dr. Dobb's Journal of Software Tools*, 2000.

[88] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 25 :1090–1098, 2012.

[89] D. Ciresan, U. Meier, J. Masci, and J. Schmidhuber. Multi-column deep neural network for traffic sign classification. *In Proc. Vision, Image, and Signal Processing*, 32 :333–338, 2012.

[90] S. C. Turaga and al. Convolutional networks can learn to generate affinity graphs for image segmentation. *Neural Comput.*, 22 :511–538, 2010.

[91] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf. Deepface : closing the gap to human-level performance in face verification. *In Proc. Conference on Computer Vision and Pattern Recognition*, pages 1701–1708, 2014.

[92] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. *The IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–9, 06 2015.

[93] Jonathan Tompson, Ross Goroshin, Arjun Jain, Yann LeCun, and Christoph Bregler. Efficient object localization using convolutional networks. *CoRR*, abs/1411.4280, 2014.

[94] Silver David, Schrittwieser Julian, Simonyan Karen, Antonoglou Ioannis, Huang Aja, Guez Arthur, Hubert Thomas, Baker Lucas, Lai Matthew, Bolton Adrian, Chen Yutian, Lillicrap Timothy, Hui Fan, Sifre Laurent, van den Driessche George, Graepel Thore, and Hassabis Demis. Mastering the game of go without human knowledge. *Nature*, pages 354–359, 2017.

[95] Siddhartha Chandra, Nicolas Usunier, and Kokkinos Iasonas. Dense and low-rank gaussian crfs using deep embeddings. *The IEEE International Conference on Computer Vision*, pages 5103–5112, 2017.

[96] Jing Liao, Yuan Yao, Lu Yuan, Gang Hua, and Sing Bing Kang. Visual attribute transfer through deep image analogy. *The IEEE Conference on Computer Vision and Pattern Recognition*, 2017.

[97] Pierre Garrigues, Sachin Sudhakar Farfade, Hamid Izadinia, Kofi Boakye, and Yannis Kalantidis. Tag prediction at flickr : a view from the darkroom.

Proceedings of Workshop on Large Scale Computer Vision Systems in NIPS, 2016.

[98] Huang Haibin, Evangelos Kalogerakis, Siddhartha Chaudhuri, Duygu Ceylan, Vladimir Kim, and Ersin Yumer. Learning local shape descriptors from part correspondences with multiview convolutional networks. *ACM Transactions on Graphics*, 37 :1–14, 11 2017.

[99] Yann LeCun, Y Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521 :436–44, 05 2015.

[100] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. Imagenet large scale visual recognition challenge. *International Journal of Computer Vision (IJCV)*, 115 :211–252, 2015.

[101] Pytorch. `https://github.com/pytorch/pytorch`.

Teddy DEBROUTELLE

Détection et classification de décors gravés sur des céramiques anciennes par analyse d'images

Le projet ARCADIA vise à développer une méthode automatique d'analyse des décors sur des tessons de céramique réalisés à la molette pour faciliter l'interprétation de ce patrimoine archéologique. Cette automatisation doit remplacer la procédure manuelle effectuée par l'archéologue, devenue trop fastidieuse avec l'augmentation du corpus (38000 tessons). L'objectif in fine est de réussir à associer automatiquement les décors à la molette du potier qui les a créés. Dans ce contexte, nous avons développé une chaine complète depuis la numérisation des tessons jusqu'à la classification automatique des décors selon leur style de motifs (carré, losange, chevrons, oves, etc). Les travaux présentés proposent plusieurs contributions mettant en œuvre des méthodes d'analyse d'images et d'apprentissage automatique. A partir du nuage de points 3D, une carte des profondeurs est obtenue. Une méthode originale de détection automatique de la région saillante focalisée sur le décor est proposée. Ensuite les décors sont caractérisés pour effectuer leur classification. Un nouveau descripteur, appelé Blob-SIFT, est proposé pour collecter les signatures seulement dans les zones pertinentes du décor. Cette approche adaptée à chaque décor, permet à la fois de réduire considérablement la masse de données et d'améliorer les performances de classification. Nous proposons également une approche apprentissage profond, puis, une approche hybride combinant les vecteurs de caractéristiques locales extraites par Blob-SIFT et la caractérisation globale du décor fournie par l'apprentissage profond qui améliore encore les performances de classification.

Mots clés : analyse d'images, classification, reconnaissance de motifs gravés, céramiques médiévales

Extraction and classification of engraved ceramic sherds by image analysis

The ARCADIA project aims to develop an automatic method for analyzing engraved decorations on ceramic sherds to facilitate the interpretation of this archaeological heritage. It is to replace the manual and tedious procedure carried out by the archaeologist since the corpus increased to more 38000 sherds. The ultimate goal is to grouping all the decorations created with the same wheel by a poter. We developped a complete chain from the 3D scanning of the sherd to the automatic classification of the decorations according to their style (diamonds, square, chevrons, oves, etc). In this context, several contributions are proposed implementing methods of image analysis and machine learning. From the 3D point cloud, a depth map is extracted and an original method is applied to automatically detect the salient region centered onto the decoration. Then, a new descriptor, called Blob-SIFT, is proposed to collect signatures only in the relevant areas and characterize the decoration to perform the classification. This approach adapted to each sherd, allows both to reduce significantly the mass of data and improve classification rates. We also use deep learning, and propose an hybrid approach combining local features extracted by Blob-SIFT with global features provided by deep learning to increase the classification performance..

Keywords : image analysis, machine learning, engraved decoration, archaelogical ceramics

PRISME, 12 rue de Blois 45067 Orléans