UNIVERSITÉ DU QUÉBEC EN OUTAOUAIS

DESCRIPTEURS DE TEXTURES D'IMAGES AVEC INVARIANCE À LA ROTATION ET APPROCHE MULTI-ÉCHELLES POUR LA RECHERCHE D'IMAGES PAR LE CONTENU

MÉMOIRE PRÉSENTÉ COMME EXIGENCE PARTIELLE DU PROGRAMME DE MAÎTRISE EN SCIENCES ET TECHNOLOGIES DE L'INFORMATION

Présenté par Ayawo Désiré DANDJI

Été 2025

Jury d’évaluation

Président du Jury : Dr. Etienne Gael Tajeuna

Membre du Jury : Dr. Etienne St-Onge

Directrice de recherche : Dr. Nadia Baaziz

DÉDICACE

À ma mère et à mon frère,
pour leur amour inconditionnel, leurs sacrifices silencieux et leur présence indéfectible.

À ma conjointe,
dont la patience, l'amour et la douceur ont accompagné chaque étape de cette aventure.
Enceinte de notre fils à venir, elle incarne à mes yeux la force tranquille et la promesse de demain.

À mes enseignants bienveillants,
qui, par leur exigence éclairée et leur engagement, ont nourri en moi le goût du savoir et le sens de l'excellence.

À tous mes amis,
dont l'amitié fidèle et les encouragements ont jalonné ce parcours avec bienveillance,
recevez aussi ma profonde reconnaissance.

REMERCIEMENTS

Je tiens, tout d'abord, à exprimer ma profonde gratitude à toutes les personnes qui ont contribué, de près ou de loin, à la réalisation de ce mémoire.

Mes remerciements les plus sincères s'adressent particulièrement à ma directrice de recherche à l'Université du Québec en Outaouais (UQO), Madame la professeure Nadia Baaziz, pour son accompagnement rigoureux, ses conseils avisés, sa disponibilité constante et la qualité de son encadrement tout au long de ce projet. Son exigence bienveillante et son soutien ont été déterminants dans l'aboutissement de ce travail.

Je tiens également à remercier chaleureusement les membres du jury pour l'attention portée à ce mémoire, ainsi que l'ensemble du corps professoral de l'UQO, dont l'enseignement et l'accompagnement tout au long de mon parcours ont largement contribué à mon développement académique et scientifique.

Table des matières

Liste des figures

Liste des tableaux

Liste des abréviations, sigles et acronymes

CBIR : Content-Based Image Retrieval (Recherche d'images par le contenu)

CNN : Convolutional Neural Network (Réseau de neurones convolutif)

LBP : Local Binary Patterns (Motifs binaires locaux)

RI-LBP : Rotation Invariant Local Binary Patterns (Motifs binaires locaux invariants à la rotation)

DWT : Discrete Wavelet Transform (Transformée en ondelettes discrètes)

SWT : Stationary Wavelet Transform (Transformée en ondelettes stationnaires).

RGB: Red, Green, Blue (Rouge, Vert, Bleu)

SVM : Support Vector Machine (Machine à vecteurs de support)

k-NN: k-Nearest Neighbors (k plus proches voisins)

ZMs : Zernike Moments (Moments de Zernike)

Gabor : Décomposition par des filtres de Gabor

TPU (Tensor Processing Unit) : accélérateur Google spécialisé dans les opérations sur tenseurs

GPU (Graphics Processing Unit) : unité de calcul parallèle pour l'entraînement de modèles

ResNet50 : Residual Network-50 (Réseau à connexions résiduelles de 50 couches)

VGG16: Visual Geometry Group-16 (Réseau à 16 couches du Visual Geometry Group)

RÉSUMÉ

À l'ère du numérique, la quantité croissante d'images produites et diffusées engendre des bases de données visuelles de plus en plus volumineuses, rendant nécessaire le développement de méthodes efficaces pour leur gestion et leur exploration. Dans ce contexte, les systèmes de recherche d'images par le contenu (CBIR) reposent sur l'extraction de caractéristiques visuelles pertinentes permettant d'identifier, comparer et retrouver des images similaires. Parmi ces caractéristiques, la texture joue un rôle fondamental, notamment dans les domaines où l'analyse visuelle est primordiale. Toutefois, la robustesse des descripteurs de texture peut être affectée par des variations géométriques des images, notamment les rotations.

Ce mémoire propose le développement d'une approche hybride d'extraction de caractéristiques texturales, combinant les motifs binaires locaux (RI-LBP), réputés pour leur invariance à la rotation, avec la transformée en ondelettes stationnaires (SWT), permettant une analyse multi-échelle tout en préservant la structure spatiale des images. Cette combinaison vise à produire des descripteurs compacts, discriminants et robustes à des variations rotationnelles, adaptés aux systèmes CBIR conventionnels.

En parallèle, une approche basée sur le *transfer learning*, exploitant les couches intermédiaires de réseaux de neurones convolutionnels pré-entraînés (VGG16 et ResNet50), a été mise en œuvre afin d'extraire automatiquement des représentations hiérarchiques à partir d'images soumises à différentes orientations, puis d'appliquer une agrégation multi-angle. Une analyse comparative a été menée entre cette approche profonde et la méthode hybride proposée.

Les expérimentations ont été réalisées sur trois bases de données de référence en analyse de textures, VisTex, Outex et Kylberg, en tenant compte de différents formats d'image (niveaux de gris, RGB) et de paramètres clés (nombre de voisins, rayons, niveaux de décomposition, activations, etc.). Les résultats obtenus montrent que l'approche hybride LBP+SWT offre un bon compromis entre performance, compacité et efficacité computationnelle, la rendant particulièrement adaptée aux contextes contraints en ressources. De son côté, l'approche par transfer *learning* avec agrégation multi-angle atteint des taux de rappel supérieurs, au prix d'un coût en calcul et en mémoire nettement plus élevé. Ces résultats illustrent des compromis méthodologiques importants et ouvrent la voie à des stratégies de fusion ou d'adaptation selon les exigences des applications.

ABSTRACT

In the digital age, the increasing volume of images being produced and shared is leading to ever-growing visual databases, making it necessary to develop efficient methods for their management and exploration. In this context, Content-Based Image Retrieval (CBIR) systems rely on the extraction of relevant visual features to identify, compare, and retrieve similar images. Among these features, texture plays a fundamental role, particularly in domains where visual analysis is essential. However, the robustness of texture descriptors can be affected by geometric variations in images, notably rotations.

This thesis proposes the development of a hybrid approach for texture feature extraction, combining rotation-invariant Local Binary Patterns (RI-LBP), known for their robustness to rotation, with the Stationary Wavelet Transform (SWT), which enables multi-scale analysis while preserving the spatial structure of images. This combination aims to produce compact, discriminative, and rotation-robust descriptors suitable for conventional CBIR systems.

In parallel, a transfer learning-based approach was implemented, leveraging the intermediate layers of pre-trained convolutional neural networks (VGG16 and ResNet50) to automatically extract hierarchical representations from images subjected to different orientations, followed by a multi-angle aggregation strategy. A comparative analysis was conducted between this deep approach and the proposed hybrid method.

Experiments were carried out on three benchmark texture databases VisTex, Outex, and Kylberg considering different image formats (grayscale, RGB) and key parameters (number of neighbors, radii, decomposition levels, activations, etc.). The results show that the hybrid LBP+SWT approach offers a good trade-off between performance, compactness, and computational efficiency, making it particularly well-suited for resource-constrained scenarios. On the other hand, the transfer learning approach with multi-angle aggregation achieves higher recall rates, at the cost of significantly higher computational and memory demands. These results highlight important methodological trade-offs and pave the way for fusion or adaptation strategies depending on application requirements.

INTRODUCTION

Au cours des deux dernières décennies, le domaine de l'imagerie numérique a connu une expansion spectaculaire, portée par la démocratisation des smartphones, l'essor des appareils photo haute résolution et la prolifération des plateformes de partage visuel telles que Facebook, Instagram ou TikTok. Des milliards d'images sont désormais générées et partagées quotidiennement, formant des bases de données visuelles massives, dont la gestion et l'exploitation exigent des outils de recherche automatisés, précis et robustes.

Les premières approches de recherche d'images reposaient essentiellement sur des annotations manuelles ou des mots-clés associés aux contenus visuels. Toutefois, cette méthode, subjective et tributaire du langage ou des conventions culturelles, a montré ses limites en termes de cohérence, d'évolutivité et de pertinence [1], [2]. Face à ces contraintes, la communauté scientifique a introduit, dès les années 1990, les systèmes de recherche d'images par le contenu (*Content-Based Image Retrieval*, ou CBIR), capables d'extraire et d'analyser automatiquement des caractéristiques visuelles (couleur, forme, texture) pour retrouver les images pertinentes sans intervention humaine directe.

Parmi ces caractéristiques, la texture constitue un attribut fondamental, en particulier pour la classification d'images naturelles, biomédicales ou industrielles. Toutefois, les descripteurs texturaux classiques restent souvent sensibles aux transformations géométriques, notamment les rotations et les variations d'échelle, ce qui compromet leur efficacité dans des contextes réalistes de recherche.

Dans ce mémoire, nous nous intéressons à la conception et à l'évaluation de descripteurs de texture robustes, adaptés aux systèmes CBIR modernes, avec deux propriétés fondamentales :

- Une invariance à la rotation, essentielle pour les textures apparaissant sous différentes orientations ;
- Une analyse multi-échelle, permettant de capturer à la fois les structures globales et les détails fins.

Notre approche repose sur une méthode hybride combinant les motifs binaires locaux invariants à la rotation (RI-LBP) réputés pour leur robustesse aux rotations locales et la transformée en ondelettes stationnaires (SWT), qui permet une décomposition multi-résolutions sans perte d'information spatiale. Par ailleurs, pour enrichir cette étude et situer les performances de notre méthode dans le contexte des techniques récentes, nous intégrons également une approche basée sur le *transfer learning*, en exploitant les couches intermédiaires des réseaux convolutifs pré-entraînés VGG16 et ResNet50, associés à une stratégie d'agrégation multi-orientations.

La première section de ce mémoire expose la problématique abordée ainsi que les enjeux spécifiques liés à la robustesse des descripteurs texturaux dans les systèmes CBIR. La deuxième section présente un état de l'art détaillé des approches existantes, à la fois traditionnelles, et celles fondées sur l'apprentissage profond, avec un accent particulier sur les techniques visant à garantir l'invariance à la rotation. La troisième section précise les objectifs de la recherche et décrit en

détail la méthodologie adoptée, incluant les bases de données utilisées, les descripteurs extraits et les métriques de performance retenues. Les résultats obtenus, accompagnés d'une analyse comparative rigoureuse entre les approches *hand-crafted* et celles fondées sur le transfert d'apprentissage (*transfer learning*), sont présentés dans la section 4. Enfin, la section 5 conclut le mémoire en résumant les principales contributions et en proposant des pistes d'extension pour les travaux futurs.

1. PROBLÈME

La texture constitue l'une des caractéristiques visuelles fondamentales utilisées dans de nombreuses tâches de traitement et d'analyse d'images, notamment dans les domaines de la médecine, de la surveillance, de la vision industrielle ou des bases de données multimédia. Elle joue un rôle déterminant dans la reconnaissance d'objets, la classification d'images et la recherche d'images par le contenu. Toutefois, la capacité des systèmes à analyser efficacement les textures se heurte généralement à deux défis majeurs : garantir l'invariance aux rotations des images et réaliser une analyse multi-échelle pertinente [1], [2].

L'invariance à la rotation signifie que les descripteurs extraits d'une image doivent rester stables, quelle que soit l'orientation de cette dernière. Cette propriété est cruciale dans les environnements réels, où les objets texturés peuvent apparaître sous divers angles. Parallèlement, une approche multi-échelle permet de capturer à la fois les structures globales et les détails fins d'une texture, renforçant ainsi la robustesse et la capacité discriminante du système [3].

Certaines approches classiques, telles que les Local Binary Patterns (LBP) ou les Histogrammes de Gradients Orientés (HOG), ont été largement utilisées pour la description des textures, mais leurs performances sont souvent limitées face à des transformations géométriques complexes ou à des variations d'échelle [4]. D'autres approches, plus avancées, s'appuient sur des transformées multi-résolutions ou des réseaux neuronaux convolutifs (CNN) pour enrichir la représentation des textures et accroître l'invariance à la rotation [5]. Cependant, ces techniques nécessitent encore des ajustements méthodologiques et une meilleure compréhension de leurs performances en contexte réel, notamment lorsqu'elles sont intégrées dans des systèmes CBIR à grande échelle [6].

Dans ce contexte, ce mémoire vise à concevoir une méthodologie d'extraction de caractéristiques texturales à la fois invariante à la rotation et multi-échelle, adaptée aux systèmes de recherche d'images par le contenu (CBIR). L'objectif est de concevoir un système CBIR capable de récupérer efficacement des images similaires en termes de texture, quelle que soit leur orientation.

Enfin, ce travail vise ainsi à contribuer à l'avancement de l'état de l'art des descripteurs texturaux dans le cadre des systèmes CBIR, et à proposer des solutions efficaces et généralisables pour la recherche d'images robustes aux variations géométriques, avec des applications concrètes dans divers domaines exigeant précision et rapidité de traitement [8].

2. ÉTAT DE L'ART

La recherche d'images par le contenu (CBIR) a émergé comme une solution pour surmonter les limitations des systèmes d'indexation d'images basés sur des annotations textuelles. Ces systèmes reposent sur la subjectivité des utilisateurs et la langue, rendant l'indexation de vastes bases de données d'images imprécise et inefficace. Les systèmes CBIR exploitent directement les caractéristiques visuelles, telles que la couleur, la texture et la forme, pour automatiser la recherche d'images [1].

L'analyse de la texture est un aspect fondamental des systèmes CBIR, permettant de différencier des images basées sur des motifs visuels réguliers ou répétitifs. Elle joue un rôle essentiel dans plusieurs domaines. Par exemple, en imagerie médicale, l'analyse de la texture est utilisée pour détecter des anomalies tissulaires, comme dans la classification des tissus cancéreux à partir de mammographies ou d'imagerie par résonnance magnétique. Dans le domaine de la télédétection, l'analyse de la texture permet de différentier les types de sols, de forêts ou de cultures dans des images satellites [3]. Dans l'e-commerce, les systèmes CBIR utilisent l'analyse de la texture pour recommander des produits similaires en fonction des caractéristiques visuelles des tissus, vêtements ou accessoires, améliorant ainsi les fonctionnalités de recherche visuelle pour les consommateurs. Cela permet aux utilisateurs de trouver des produits ayant des textures similaires à partir de simples photos ou de captures d'écran [4]. Enfin, dans l'industrie de la qualité, la texture est exploitée pour identifier des défauts sur des surfaces de produits manufacturés, tels que des textiles ou des pièces métalliques [5].

Dans cette section, nous présentons les concepts clés de la recherche d'images par le contenu ainsi que les développements récents en matière d'invariance à la rotation et d'analyse multi-échelles.

2.1. Architecture d'un système CBIR

Un système de recherche d'images par le contenu (CBIR) conventionnel repose sur des descripteurs d'images dont l'extraction est manuellement conçue (*handcrafted*). L'architecture d'un tel CBIR se divise en deux modules : un module hors ligne et un module en ligne. Le module en ligne est dédié à la phase de la recherche d'images similaires ou correspondantes à une image requête. Cette phase commence par l'extraction des caractéristiques visuelles discriminantes de l'image soumise, lesquelles sont ensuite comparées à celles des images de la base de données, extraites préalablement lors de la phase hors ligne [1].

Les résultats de cette comparaison sont classés selon leur similitude avec l'image requête. Ce processus de classement, appelé "ranking", est essentiel pour améliorer l'expérience utilisateur. Le classement est souvent affiché dans une liste ordonnée, montrant les résultats les plus pertinents en premier (généralement les "Top N" résultats). Par exemple, si N = 10, le système renverra à l'utilisateur les 10 images les plus pertinentes par rapport à l'image requête. Le diagramme bloc d'un CBIR conventionnel est montré sur la figure 1 ci-après :

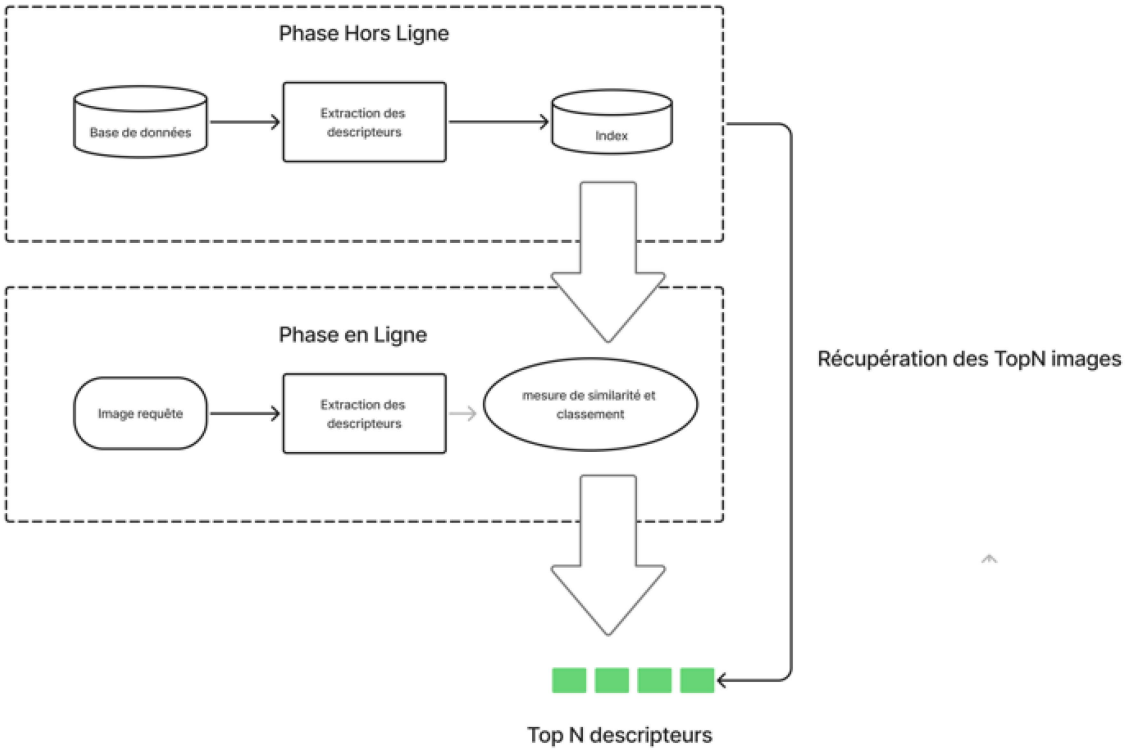


Figure 1 : Schéma fonctionnel du processus de recherche d'images par contenu (CBIR) conventionnel, illustrant les phases hors ligne (indexation) et en ligne (recherche et classement).

Phase hors ligne :

- Prétraitement **:** Les images de la base de données sont préparées à travers des opérations telles que la conversion de format, la normalisation et la suppression du bruit [3] ;
- Extraction des caractéristiques **:** Des caractéristiques visuelles (couleur, texture, forme) sont extraits pour créer des vecteurs descripteurs représentant chaque image [1] ;
- Indexation **:** Les descripteurs extraits sont indexés dans une structure de données pour permettre une recherche rapide. Des structures telles que les arbres K-dimensionnels et les index inversés peuvent être utilisées pour améliorer la performance [4].

Phase en ligne :

- Recherche et comparaison **:** Lorsque l'utilisateur soumet une image requête, ses caractéristiques sont extraites et comparées aux vecteurs indexés de la base de données à l'aide d'une mesure de similarité, telles que la distance euclidienne ou cosinus [5] ;

- Récupération et classement **:** Les images les plus similaires sont récupérées et classées en fonction de leur score de similarité. Les utilisateurs peuvent affiner les résultats via un feedback interactif [6] ;
- Evaluation des performances du système : L'évaluation d'un système CBIR repose sur des métriques telles que :
  - Précision : le ratio d'images pertinentes parmi celles récupérées ;
  - Rappel : la capacité à retrouver toutes les images pertinentes dans la base ;
  - F-Score : une moyenne harmonique entre précision et rappel, indiquant un équilibre entre les deux.

Ces métriques permettent de comparer la performance des systèmes CBIR sur des bases de données variées et avec différentes configurations de modèles [10].

2.2. CBIR fondé sur l'apprentissage supervisé classique

L'intégration de l'apprentissage supervisé dans les systèmes CBIR repose sur une architecture semblable à celles des approches conventionnelles, combinant des phases hors ligne (extraction de caractéristiques et construction du modèle) et en ligne (recherche et récupération des images).

Dans ce cadre, les méthodes d'extraction de descripteurs d'images sont conçues manuellement (*handcrafted*), comme dans les approches conventionnelles. Ces descripteurs sont ensuite utilisés comme entrées pour des algorithmes d'apprentissage supervisé classiques (tels que les SVM, k-NN ou les arbres de décision), capables d'apprendre à partir des descripteurs annotés.

Ce paradigme exploite la capacité des modèles supervisés classiques et permet au système d'apprendre des représentations discriminantes et de classer les images selon des catégories définies, plutôt que de se baser uniquement sur des mesures de similarités génériques. Les deux phases de cette architecture (voir la figure 2) sont décrites ci-après :

- *Phase hors ligne (entraînement)*

Cette étape comprend la collecte d'un ensemble représentatif d'images, sa division en ensembles d'entraînement et de validation, le prétraitement (redimensionnement, normalisation, conversion), l'extraction des descripteurs (caractéristiques de texture, couleur, forme), l'indexation efficace de ces caractéristiques (par exemple avec PCA) et l'entraînement du modèle d'apprentissage supervisé (SVM, k-NN) en optimisant les paramètres par validation croisée pour éviter le sur-apprentissage.

- *Phase en ligne (recherche et récupération)*

Lorsqu'une nouvelle image requête est introduite, ses caractéristiques sont extraites, et le modèle entraîné prédit sa catégorie. Une recherche rapide est effectuée dans l'index des caractéristiques préalablement construit afin d'identifier et de récupérer les images de la même catégorie, garantissant ainsi une réponse rapide et pertinente. Toutefois, cette approche demeure limitée par la nécessité de disposer d'une base de données annotée

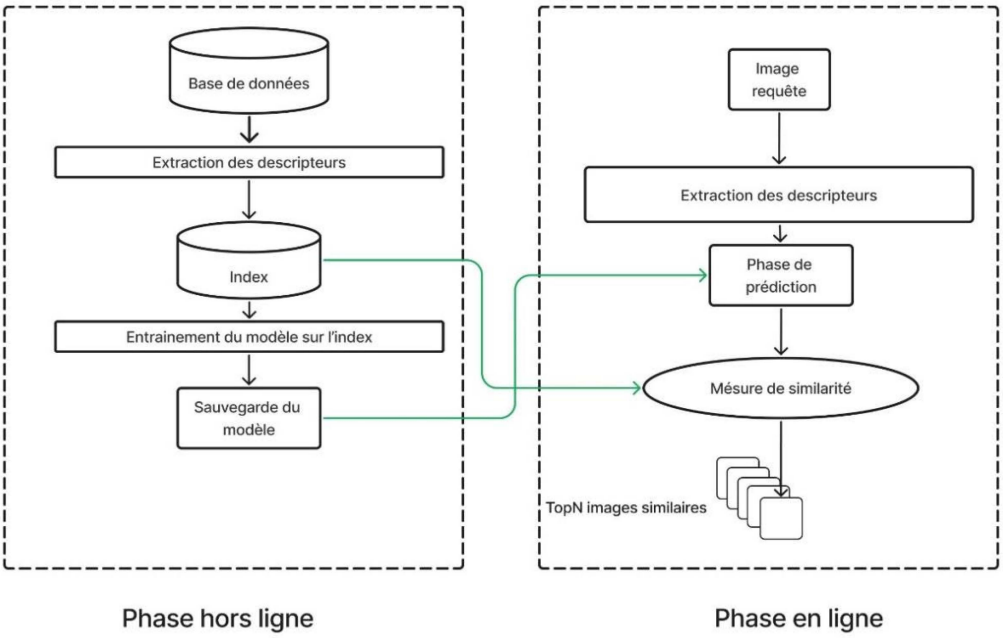


Figure 2 : Schéma fonctionnel du processus de recherche d'images par contenu (CBIR) intégrant l'apprentissage supervisé.

Les flèches vertes indiquent le transfert des informations de la phase hors ligne vers la phase en ligne : l'index des descripteurs sert de référence pour la comparaison, tandis que le modèle entraîné est chargé pour effectuer la prédiction et la mesure de similarité lors de la recherche d'images similaires.

2.3. CBIR fondé sur l'apprentissage profond

L'intégration de l'apprentissage profond dans les architectures de systèmes CBIR permet d'apprendre automatiquement des représentations discriminantes des images. Les réseaux de neurones convolutifs (CNN) constituent actuellement une approche répandue, mais d'autres techniques d'apprentissage profonds peuvent également être exploitées (auto-encodeurs, transformers, etc.). Contrairement aux deux approches CBIR précédentes, les caractéristiques ne sont plus extraites par des méthodes développées manuellement (*hand-crafted*), mais directement apprises à partir de grandes bases d'images annotées.

Tout comme les architectures précédentes, le CBIR fondé sur un réseau de neurones convolutif comprend une phase hors ligne (entraînement et indexation) et en ligne (recherche et récupération) dont voici les principales étapes :

- *Phase hors ligne (entraînement et indexation)*

Cette phase englobe la collecte et la préparation des images (redimensionnement, normalisation, conversion d'espaces couleur), leur étiquetage en classes spécifiques, et l'entraînement d'un réseau de neurones convolutif (CNN) supervisé. Le CNN est entraîné sur des données étiquetées pour optimiser les paramètres du réseau afin d'extraire automatiquement des caractéristiques discriminantes. Ces caractéristiques sont ensuite indexées pour une récupération efficace.

- *Phase en ligne (recherche et récupération)*

Lorsqu'une image requête est soumise, ses caractéristiques sont extraites via le modèle CNN entraîné. Ces caractéristiques sont comparées aux vecteurs descripteurs indexés en utilisant des métriques de similarité appropriées pour récupérer rapidement les images les plus similaires.

Cette approche basée sur l'apprentissage profond, bien que puissante, présente des défis majeurs : nécessité d'importants volumes de données étiquetées pour l'entraînement, de ressources de calcul importantes, et risque élevé de sur-apprentissage en cas de données limitées ou mal annotées.

Pour pallier la nécessité de disposer d'une vaste base d'images annotées, une variante avantageuse de cette approche consiste à adopter l'apprentissage par transfert (*transfer learning*), en exploitant des réseaux profonds pré-entraînés (tels que ResNet-50 ou VGG-16) comme extracteurs de caractéristiques. Les images de la base cible sont alors passées dans ces modèles, dont les couches convolutionnelles sont figées. Cette stratégie réduit drastiquement le temps d'entraînement et la quantité de données annotées requises, tout en conservant une grande richesse sémantique dans les descripteurs extraits. La recherche d'images peut ensuite s'effectuer à l'aide de l'algorithme des k plus proches voisins (k-NN) et la distance euclidienne, comme illustré dans l'exemple de la figure 3 ci-après.

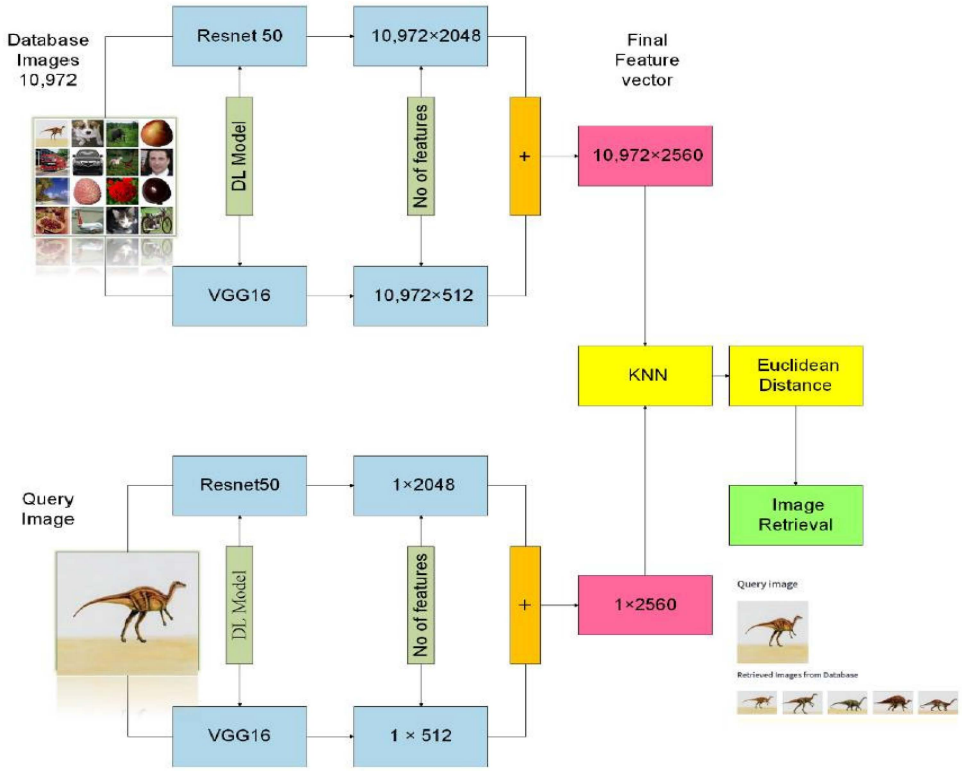


Figure 3: Schéma bloc d'un système de recherche d'images par contenu (CBIR) basé sur le transfert d'apprentissage, combinant les caractéristiques profondes extraites de ResNet50 et VGG16 et la similarité mesurée par la distance euclidienne à l'aide du classifieur KNN

2.4. Approches d'extraction de descripteurs de texture

La texture est une caractéristique visuelle essentielle qui se retrouve sur presque toutes les surfaces, qu'elles soient naturelles, comme les feuilles ou le bois, ou artificielles, comme les tissus ou les murs. Elle joue un rôle fondamental dans l'analyse d'images, permettant de capturer les caractéristiques texturales qui ne peuvent être facilement perçues par les descripteurs basés uniquement sur la couleur ou la forme. Les textures apportent une dimension supplémentaire dans la différenciation d'objets ou de régions dans une image, en particulier lorsque les motifs structuraux sont réguliers ou présentent des répétitions visibles.

Au fil des années, de nombreuses méthodes ont été proposées pour extraire ces caractéristiques, chacune offrant des avantages spécifiques en fonction du type de texture à analyser et de l'application visée. Ces méthodes varient selon la manière dont elles modélisent la texture, allant des approches basées sur les statistiques locales aux approches plus complexes qui exploitent des transformations mathématiques ou des réseaux de neurones profonds. Chaque approche cherche à fournir une représentation pertinente des variations texturales dans une image, afin d'améliorer les performances dans des tâches telles que la classification, la segmentation ou la récupération d'images par le contenu.

Dans cette section, nous allons explorer et classer les principales approches d'extraction de descripteurs de texture, en détaillant leurs principes fondamentaux ainsi que leurs avantages et inconvénients dans différentes applications pratiques selon la taxonomie proposée dans [16].

2.4.1. Approches statistiques

Les approches statistiques reposent sur l'analyse des propriétés statistiques de la distribution spatiale des niveaux de gris dans une image. Elles mesurent des relations entre les pixels, telles que la fréquence de cooccurrence des valeurs d'intensité et la répartition des motifs dans l'image. L'un des exemples les plus courants est la matrice de cooccurrence des niveaux de gris GLCM (Gray-Level Co-Occurrence matrix), qui capture les relations entre les pixels à différentes distances et orientations pour décrire des caractéristiques comme l'énergie, l'entropie et la corrélation. Un autre exemple est la méthode des motifs binaires locaux LBP (Local Binary Patterns), qui encode les relations de voisinage des pixels sous forme de motifs binaires, afin de capturer des structures locales comme les bords et les coins dans les images de textures [2].

2.4.2. Approches structurales

Ces méthodes décomposent la texture en éléments de base appelés primitives et analysent leur arrangement spatial. Elles sont particulièrement adaptées aux textures régulières et répétitives. Ces approches permettent d'identifier les primitives et de définir des règles d'agencement pour caractériser la texture globale de l'image. Elles sont souvent utilisées pour les analyses où des motifs géométriques bien définis sont présents, telles que la détection de bords et la reconnaissance de formes répétées dans les images.

2.4.3. Approches basées sur des transformations

Les images sont transformées dans un espace fréquentiel ou d'échelle afin d'analyser la texture à différentes échelles et fréquences [13]. Ces méthodes permettent d'extraire des informations sur la texture à plusieurs niveaux d'échelle, en décomposant les images en sous-bandes fréquentielles. Parmi les approches les plus courantes figurent les variantes de la transformée en ondelettes, qui permettent de capturer à la fois des détails fins et grossiers des textures à différentes échelles. Ces approches trouvent de nombreuses applications en classification et en segmentation d'images [5].

2.4.4. Approches basées sur des modèles

Ces approches modélisent la texture à l'aide de modèles mathématiques comme les champs aléatoires ou les modèles autorégressifs. Ces modèles capturent les relations locales entre les pixels

et sont souvent utilisés pour caractériser des textures complexes ou aléatoires [16]. Les modèles de champs aléatoires sont largement utilisés pour représenter des textures stochastiques, en prenant en compte les relations spatiales entre pixels.

2.4.5. Approches basées sur les graphes

Les approches basées sur les graphes représentent l'image sous forme de graphe, où les pixels sont des nœuds et les relations entre les pixels sont représentées par des arêtes. Ces méthodes permettent d'analyser la connectivité et la structure globale de l'image en utilisant des mesures de théorie des graphes, comme la connectivité locale ou la distance entre les pixels [16]. Les structures locales de graphes sont utilisées pour représenter les connexions entre les pixels et pour analyser les propriétés globales de la texture.

2.4.6. Approches basées sur l'entropie

Ces méthodes mesurent la quantité d'information ou le degré de désordre dans une texture à l'aide de calculs d'entropie. Elles sont utiles pour caractériser la complexité et l'uniformité d'une texture, en se basant sur des concepts de la théorie de l'information [37]. Par exemple, l'entropie de Shannon est utilisée pour quantifier la distribution des motifs dans les images et caractériser leur régularité ou leur complexité.

2.4.7. Approches basées sur l'apprentissage profond

Les approches fondées sur l'apprentissage profond (*deep learning*) exploitent différentes architectures de réseaux de neurones pour extraire automatiquement des caractéristiques texturales riches, en capturant des informations multi-niveaux dans les représentations texturales [9], directement à partir des données visuelles, sans recourir à des descripteurs prédéfinis manuellement (*hand-crafted*).
Parmi celles-ci, les réseaux de neurones convolutifs (CNN) se distinguent par leur capacité à apprendre des descripteurs discriminants grâce à leurs couches de convolution et de pooling, qui hiérarchisent l'extraction de détails locaux jusqu'aux structures globales de l'image. Par exemple, la méthode Local to Global (L2G) combine efficacement ces deux niveaux d'information pour renforcer la robustesse des systèmes CBIR [17], tandis que l'utilisation de filtres invariants à la rotation et de pooling multi-échelles a permis d'améliorer encore plus la précision des requêtes [18], notamment sur des tâches de reconnaissance de textures [38].

Toutefois, si ces architectures profondes offrent des performances remarquables, leur utilisation s'accompagne souvent d'un coût computationnel élevé et d'un besoin important en données annotées, notamment pour éviter le surapprentissage. En effet, l'entraînement complet de réseaux tels que ResNet nécessite des ressources matérielles importantes (GPU/TPU) et des temps de

calcul prolongés, ce qui peut constituer un obstacle dans les contextes de bases de données de taille limitée ou des applications en temps réel.

Pour surmonter ces contraintes, l'apprentissage par transfert (*transfer learning*) constitue une approche efficace. Il consiste à réutiliser un réseau convolutif déjà entraîné sur une base d'images génériques de grande envergure, dans le cadre d'une tâche de classification à large échelle. Les couches–initiales ou intermédiaires du modèle sont alors conservées pour l'extraction de descripteurs visuels riches et généraux, tandis que les couches supérieures ou finales peuvent, si nécessaire, être affinées et ajustées aux données spécifiques de la tâche cible. Dans le cas du CBIR, cette étape d'adaptation des couches finales du réseau n'est généralement pas requise, puisque l'objectif n'est pas la classification mais la comparaison d'images fondée sur la similarité entre leurs descripteurs. Cette stratégie permet de réduire considérablement le volume de données annotées nécessaire et le coût de calcul, tout en assurant une bonne qualité de représentation de l'information visuelle [39]. Ainsi, en combinant réseaux de neurones convolutifs et apprentissage par transfert, les approches modernes en CBIR parviennent à un compromis équilibré entre précision, robustesse et efficience computationnelle.

Par ailleurs, il est à noter que cette approche de transfer learning peut être complémentaire aux méthodes hand-crafted.

En résumé, les méthodes d'extraction de descripteurs de textures dans un contexte CBIR se répartissent en deux grandes familles. La première, dite *hand-crafted*, regroupe les méthodes conçues manuellement, reposant sur des principes statistiques, structurels, basés sur les transformées, la modélisation, les graphes ou l'entropie. Ces descripteurs se distinguent par leur transparence, leur faible coût de calcul et leur capacité à bien fonctionner sur de petits jeux de données, ainsi qu'une forte interprétabilité. La seconde famille repose sur l'apprentissage profond, en particulier les réseaux de neurones convolutifs (CNN) et l'apprentissage par transfert (*transfer learning*). Ces approches permettent d'apprendre automatiquement des représentations hiérarchiques riches et discriminantes. Elles remplacent l'extraction manuelle par une modélisation automatique des caractéristiques à partir de grandes bases d'images annotées. Ce qui entraine un besoin de ressources accru et une interprétabilité des descripteurs souvent difficile et réduite. Récemment, des travaux se sont intéressés à la fusion de ces deux stratégies en combinant des descripteurs manuels et des caractéristiques apprises ouvre des perspectives prometteuses pour la conception de systèmes CBIR modernes [40], [41].

Le tableau 1 ci-contre présente une synthèse comparative des principales familles d'approches existantes, leurs avantages, leurs limites, ainsi que leurs domaines d'application typiques.

Tableau 1 : Synthèse comparative des approches d'extraction de descripteurs de texture

Famille d'approches	Principe-clé	Exemples représentatifs	Atouts	Limites	Applications
Statistiques	Relations locales d'intensité	GLCM, LBP/RI-LBP	Interprétable, Simple, léger	Sensible au bruit/rotation, dépend des paramètres	Classification, segmentation, CBIR
Structurales	Primitives + agencement spatial	Primitives, détection de bords	Adéquate pour motifs réguliers	Faible sur textures stochastiques	Inspection visuelle, motifs répétitifs
Transformations	Analyse multi-échelles/fréquences	Ondelette (DWT/SWT), Gabor, Fourier	Bonne sélectivité échelle/orientation	Coût supérieur, réglages fins requis	Classification multi-échelle, CBIR
Modèles	Modélisation probabiliste locale	MRF/CRF, AR/ARMA	Solide pour textures aléatoires	Estimation/optimisation lourdes	Synthèse, segmentation
Graphes	Image → graphe (nœuds/arêtes)	Graphlets, MST, GLCG	Capture structure globale/connectivité	Coût mémoire/temps, choix du graphe	Textures structurées, scènes
Entropie	Mesure du désordre/complexité	Entropies de Shannon/Rényi	Ultraléger, complémentaire	Seul peu discriminant	Qualité/contraste, pré-filtrage
Apprentissage profond	CNN/transformers + transfer learning	VGG16, ResNet50, L2G	Très discriminant, robuste	Coût mémoire/temps très élevé, faible interprétabilité	CBIR, imagerie médicale, télédétection

2.5. Invariance à la rotation dans l'analyse de textures

L'invariance à la rotation est une caractéristique clé dans les systèmes de recherche d'images par le contenu (CBIR), la classification et la reconnaissance de textures, car les motifs texturaux dans les images peuvent apparaître sous différents angles en fonction de l'orientation de l'objet capturé ou de la caméra utilisée pour l'acquisition de l'image. Garantir que les descripteurs de texture soient invariants à ces rotations est essentiel pour améliorer la robustesse et la précision des systèmes de recherche. Les textures, par leur nature géométrique et répétitive, sont souvent soumises à des variations d'orientation. Pour surmonter ce défi, des méthodes spécifiques ont été développées pour garantir que les caractéristiques extraites des textures restent cohérentes, peu importe l'angle de rotation. Une des méthodes les plus populaires est l'extension des Local Binary Patterns (LBP) en Rotation Invariant Local Binary Pattern (RI-LBP), qui modifie les LBP classiques pour rendre les motifs locaux insensibles aux rotations. Cette méthode s'est avérée efficace pour capturer des structures texturales répétitives, indépendamment de leur orientation dans l'image [2].

Cependant, plusieurs autres travaux récents ont également utilisé le LBP pour garantir l'invariance à la rotation. Une méthode améliorée du LBP, appelée Multi-Scale LBP, qui intègre plusieurs résolutions pour mieux capturer les motifs à différentes échelles tout en garantissant l'invariance à la rotation, a été proposée. Cette approche a montré de bons résultats dans la classification des textures. En parallèle, Saipullah *et al.* [27] ont proposé un nouvel algorithme basé sur les différences ordonnées du voisinage (SND, Sorted Neighborhood Differences), qui élimine la

dépendance des descripteurs de texture à l'orientation des pixels voisins en triant les pixels avant d'extraire des motifs invariants. Leurs résultats montrent une amélioration significative de la robustesse aux rotations par rapport aux descripteurs LBP traditionnels.

Aussi, les moments de Zernike sont utilisés pour leur capacité à générer des descripteurs invariants à la rotation. Appliqués aux textures, ces moments permettent de capturer les caractéristiques géométriques des motifs tout en restant robustes face aux transformations angulaires [14]. Leur efficacité dans l'analyse de textures à rotation variable a été démontrée, et leur pertinence pour des textures régulières mais sujettes à des variations d'orientation est prouvée [15]. Kolte *et al.* soulignent dans [30] l'importance de la texture dans l'analyse des images rétiniennes et expliquent que l'extraction de caractéristiques via les moments de Hu permet d'obtenir des descripteurs robustes aux transformations géométriques telle que la rotation. Pour détecter la rétinopathie diabétique à partir d'images de fond d'œil, un modèle de classification prometteur est développé sur des descripteurs de texture combinant des caractéristiques de Haralick (basées sur la matrice de co-occurrence des niveaux de gris) et les sept moments invariants de Hu. Par ailleurs, dans [31], la détection automatique de diverses pathologies à partir d'images de feuilles de plante s'appuie sur l'apprentissage automatique et la classification de descripteurs obtenus en associant les moments invariants de Hu aux caractéristiques de Haralick et à l'histogramme de couleur.

En complément, des approches combinant des transformées en ondelettes et la transformation log-polaire ont été proposées pour assurer une meilleure robustesse aux rotations. Pun et Lee [25] ont introduit une méthode exploitant les signatures d'énergie d'ondelettes log-polaires pour classifier des textures sous différentes orientations et échelles. Leur approche applique d'abord une transformation log-polaire pour éliminer les effets de rotation et de mise à l'échelle, suivie d'une transformée en ondelettes adaptative permettant d'extraire des caractéristiques invariantes aux rotations et aux changements d'échelle. Cette approche a montré des taux de classification supérieurs aux méthodes traditionnelles. Une autre avancée repose sur la combinaison de la transformée en ondelettes complexes et les motifs binaires locaux pour améliorer la robustesse à la rotation. Dans [26], Yang *et al.* proposent une méthode d'extraction de caractéristiques de textures basée sur la décomposition en ondelettes complexes. Les caractéristiques locales issues de l'application du LBP sur les sous-bandes d'approximation, sont combinées aux mesures d'énergie des sous-bandes de détail dans l'espace log-polaire. Ce descripteur, testé avec l'algorithme KNN, assure une classification robuste des textures, invariante aux rotations, à l'éclairage et aux changements d'échelle, surpassant des méthodes existantes.

Farhan Riaz *et al.* proposent une méthode de classification des textures qui exploite les filtres de Gabor pour obtenir des descripteurs invariants aux rotations et aux changements d'échelle [32]. La méthode extrait d'abord des caractéristiques de texture homogène (HT) via un banc de filtres de Gabor, réorganisant les réponses en matrices dont les décalages, dus aux rotations et aux échelles, sont neutralisés par l'application de la transformée de Fourier discrète (DFT). En combinant ce vecteur de caractéristiques avec un SVM pour la classification, l'approche démontre, sur les jeux de données Brodatz, une invariance efficace aux transformations géométriques, surpassant ainsi des méthodes classiques. De plus, Chaorong Li *et al.* [28] ont recours à une variante de la décomposition de Gabor qui améliore la classification des textures en réalisant une décomposition hiérarchique du filtre de Gabor circulairement symétrique (CSGW) pour remédier

à son manque de sélectivité directionnelle. La méthode génère des sous-bandes à différentes résolutions, dont les interdépendances sont statistiquement modélisées par une copule gaussienne (avec des marges ajustées par des lois de Weibull). En combinant ces paramètres avec des mesures d'énergie, un vecteur de caractéristiques robuste est constitué, puis exploité par un SVM pour la classification. Les expérimentations sur les bases de données Outex et UIUC démontrent une invariance efficace aux rotations, surpassant ainsi des approches classiques.

Ying Liu *et al*. proposent la méthode HOG-DG (Histogramme des gradients avec Gradient dominant) qui améliore la classification des images de motifs pneumatiques en extrayant des descripteurs de texture robustes aux variations d'illumination, d'échelle et de rotation [33]. Cette méthode commence par le calcul des HOG à partir de cellules locales circulaires, garantissant ainsi la capture d'une même région indépendamment de l'angle de rotation. Ensuite, ces descripteurs sont alignés selon le gradient dominant afin de corriger la sensibilité aux rotations. En combinant ces caractéristiques avec un classifieur SVM, le vecteur de caractéristiques obtenu se révèle particulièrement efficace, comme le confirment les expérimentations sur des ensembles de données de motifs pneumatiques, et la comparaison à des approches classiques. De même, Sharma et Ghosh introduisent le descripteur de l'histogramme des magnitudes de gradients (HGM) pour l'analyse de textures [29]. Leur approche repose sur le calcul d'un histogramme des magnitudes de gradients, tout en ignorant l'information directionnelle, conférant ainsi au descripteur extrait une invariance naturelle aux rotations. De plus, HGM est de faible dimension (16 bins) et nécessite peu de calcul, ce qui le rend particulièrement adapté aux applications de classification et de segmentation d'images. Les résultats expérimentaux montrent que, seul ou combiné avec d'autres descripteurs, HGM surpasse les méthodes classiques telles que HOG et LBP.

Les approches basées sur l'apprentissage profond offrent également des mécanismes efficaces pour atteindre l'invariance à la rotation des descripteurs extraits. Lors de l'entraînement, les données sont enrichies par la génération d'images tournées selon différents angles (*data augmentation*), ce qui permet au réseau d'apprendre des représentations robustes aux variations angulaires. En outre, des méthodes spécifiques consistent à extraire les représentations intermédiaires de plusieurs versions tournées d'une même image, puis à les agréger (par exemple via une moyenne) afin de produire une représentation finale invariante aux rotations. Ces stratégies bénéficient de la richesse des représentations hiérarchiques acquises par les réseaux pré-entraînés, tout en renforçant la stabilité des descripteurs face aux transformations extrêmes. Des études récentes [39], [40] montrent que cette approche permet d'accroître significativement la robustesse des systèmes CBIR en tirant parti de la capacité des réseaux convolutifs à extraire des caractéristiques invariantes à la fois par leur architecture (pooling et filtres locaux) et par des processus d'agrégation utilisés comme post-traitements. Ainsi, la combinaison d'une stratégie d'augmentation des données par rotations et d'une agrégation des représentations intermédiaires extraites constitue une approche efficace pour atteindre l'invariance à la rotation, indispensable pour des applications telles que la classification de textures, l'analyse d'images aériennes ou l'imagerie biomédicale.

Enfin, les études confirment que l'invariance à la rotation améliore considérablement la classification et la segmentation des textures, particulièrement dans les domaines où les variations angulaires sont inévitables, comme la reconnaissance de surfaces dans les images aériennes ou les

scans médicaux [19]. Ces avancées montrent que l'invariance à la rotation demeure un élément crucial pour les systèmes CBIR spécialisés dans l'analyse des textures.

2.6. Métriques de similarité et de performances des systèmes CBIR

Les systèmes CBIR reposent sur des mesures de similarité pour comparer les images entre elles et les indicateurs de performance pour évaluer la qualité des résultats obtenus. Cette section présente les principales mesures de similarité utilisées dans les systèmes CBIR ainsi que les indicateurs de performance, avec des formules explicites et des explications détaillées.

2.6.1. Mesures de similarité

Dans un système CBIR, la similarité entre une image requête et une image de la base de données est évaluée à l'aide des descripteurs visuels, tels que les vecteurs de caractéristiques (couleur, texture, forme). Les mesures de similarité permettent de comparer ces vecteurs, les images similaires correspondant à des distances ou des valeurs de dissimilarités faibles. Voici les principales métriques de similarité utilisées dans les systèmes CBIR :

- Distance Euclidienne

La distance Euclidienne est l'une des mesures les plus couramment utilisées pour calculer la dissimilarité entre deux vecteurs de caractéristiques. Elle mesure la racine carrée de la somme des différences au carré entre les composants des deux vecteurs [1].

$$D_E(x,y) = \sqrt{(\Sigma_{i=1}^{n}(x_i - y_i)^2)}$$

- Distance de Manhattan (ou City Block)

La distance de Manhattan (ou norme L1) calcule la somme des différences absolues entre les composants des vecteurs de caractéristiques. Cette mesure est moins sensible aux grandes variations dans une seule dimension par rapport à la distance Euclidienne [2].

$$D_M(x,y) = \Sigma_{i=1}^{n}|x_i - y_i|$$

- Distance de Minkowski

La distance de Minkowski est une généralisation des distances Euclidienne et de Manhattan. Paramétrée par un entier p, elle permet d'adapter la métrique en fonction des caractéristiques spécifiques des descripteurs visuels utilisés.

$$D_p(x,y) = (\Sigma_{i=1}^{n}|x_i - y_i|^p)^{(1/p)}$$

- Distance de Mahalanobis

La distance de Mahalanobis prend en compte les corrélations entre les différentes variables des vecteurs de caractéristiques et est robuste aux différences d'échelles.

$$D_Mah(x,y) = \sqrt{\left((x-y)^T S^{-1}(x-y)\right)}$$

- Distance Cosinus

La distance cosinus mesure l'angle entre deux vecteurs de caractéristiques et est particulièrement utile pour des vecteurs de grande dimension. Elle est fréquemment utilisée dans les systèmes CBIR basés sur descripteurs texturaux ou dans des systèmes d'apprentissage automatique [3].

$$D_C(x,y) = 1 - \left((x \cdot y)/(\|x\|\|y\|)\right)$$

- Distance du Khi-Carré (ou Chi-Square)

La distance du Khi-Carré est une mesure utilisée pour évaluer la dissimilarité entre deux histogrammes ou distributions de données. Contrairement à la distance Euclidienne qui calcule la distance géométrique directe entre deux points dans un espace multidimensionnel, la distance du Khi-Carré met l'accent sur la comparaison des proportions relatives entre deux distributions. Elle est particulièrement utile lorsque les données sont sous forme d'histogrammes ou de fréquences.

$$D_{\chi^2}(x,y) = 1/2 \cdot \Sigma_{i=1}^{n}\left((x_i - y_i)^2/(x_i + y_i + \varepsilon)\right)$$

2.6.2. Indicateurs de performance des systèmes CBIR

Pour évaluer la qualité des résultats obtenus par un système CBIR, plusieurs métriques de performance sont utilisées. Avant de discuter des différents indicateurs, il est important de définir quelques termes de base qui servent à évaluer les performances du système :

- True Positives (TP) : Ce sont les images qui ont été correctement identifiées comme pertinentes. En d'autres termes, ce sont les bonnes réponses que le système a fournies ;
- False Positives (FP) : Ce sont les images qui ne sont pas pertinentes, mais que le système a incorrectement identifiées comme pertinentes. Ce sont les faux résultats positifs ;
- False Negatives (FN) : Ce sont les images pertinentes qui ont été omises par le système. Autrement dit, ce sont les bonnes réponses que le système a échappées ;
- True Negatives (TN) : Ce sont les images non pertinentes qui ont été correctement identifiées comme non pertinentes par le système.

Voici la liste des indicateurs de performance :

- Précision (*Precision*) **:** mesure le pourcentage d'images pertinentes (par rapport à la requête) parmi celles retournées par le système.

$$Précision = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}$$

- Rappel (*Recall*) **:** mesure le pourcentage d'images pertinentes retournées par rapport au nombre total d'images pertinentes disponibles dans la base de données d'images [4].

$$Rappel = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}$$

- F-score **:** est une mesure qui combine à la fois la précision et le rappel. Il est la moyenne harmonique de la précision et du rappel, offrant un équilibre entre les deux. Un score élevé indique que le système a une bonne précision et un bon rappel.

$$F_{score} = 2 * \frac{Précision * Rappel}{Précision + Rappel}$$

En résumé de cette section, les systèmes CBIR ont évolué au cours des dernières années grâce à l'introduction et l'incorporation croissante d'une variété de descripteurs visuels, et à l'apport de l'apprentissage automatique, notamment avec les réseaux neuronaux profonds. De nos jours, ces systèmes atteignent de bons niveaux de précision dans divers domaines d'applications. Cependant, des défis continuent à stimuler la recherche dans ce domaine. Parmi ceux-ci figurent la difficulté de traiter de gros volumes de données et la robustesse de ces systèmes face à la variabilité des données d'images dues aux changements d'orientations, d'éclairage ou de bruit.

3. PROPOSITION DE RECHERCHE

3.1. Objectifs

Ce projet de recherche vise à explorer l'analyse de textures d'images et l'extraction de descripteurs pertinents destinés à la recherche d'images par le contenu (CBIR). L'accent est mis sur l'invariance à la rotation et sur l'approche multi-échelle tout en recherchant une compacité accrue des descripteurs ainsi qu'un pouvoir discriminant renforcé. Les objectifs spécifiques de cette étude sont les suivants :

- Développer et valider une méthodologie d'extraction de descripteurs de textures, compacts et discriminants, intégrant une approche multi-échelle avec invariance à la rotation, en vue de son implémentation dans un système CBIR conventionnel ;
- Analyser les paramètres influençant la performance des algorithmes d'extraction et de recherche appliqués aux textures sous différentes orientations ;
- Évaluer les performances de la méthodologie proposée sur des bases de données de textures de référence, en termes d'efficacité, de robustesse et de capacité à concilier compacité et pouvoir discriminant.
- Comparer les performances de la méthode proposée avec une approche basée sur l'apprentissage profond exploitant le *transfer learning*, en étudiant les avantages et limitations respectifs en termes de précision, robustesse et coût computationnel.

3.2. Méthodologie

Un système CBIR est à développer afin de rechercher et de récupérer des images pertinentes en réponse à une requête dans une base de données d'images de textures. Ce système permettra d'indexer les images de la base en utilisant les descripteurs extraits. Lorsqu'une image requête est soumise, son descripteur est extrait, et l'index sera utilisé pour retrouver rapidement les images les plus similaires à la requête en comparant leurs descripteurs à l'aide de mesures de similarité telles que la distance euclidienne ou la distance chi-square. Dans cette étude, le module central du CBIR concerne l'extraction des caractéristiques et la construction des descripteurs de texture. Il sera basé sur une méthode qui est invariante à la rotation et multi-échelle.

L'extraction de descripteurs texturaux peut être réalisée à l'aide de méthodes exploitant à la fois les motifs locaux et globaux sur plusieurs échelles. Ces approches garantissent une représentation robuste de la texture en prenant en compte les variations d'échelle et les transformations géométriques, notamment la rotation.

- Méthodes basées sur des descripteurs locaux (statistiques) **:** permettent d'extraire des informations texturales à partir des relations entre les pixels voisins d'une image. Les variations locales de l'intensité des pixels sont analysées pour capturer des motifs caractéristiques. Leur principal atout est leur invariance à la rotation **:** elles sont conçues

pour assurer une capture cohérente des motifs locaux, quelle que soit l'orientation de l'image. Cette propriété est essentielle dans des contextes où les images sont prises sous différents angles ou lorsque l'objet d'intérêt subit une rotation. Une autre caractéristique clé est l'approche multi-échelle, qui adapte les descripteurs à différents niveaux de granularité dans l'image. Cette approche facilite l'analyse des textures à la fois grossières et fines, en capturant les détails les plus fins à des échelles plus petites, tout en prenant en compte des structures plus larges à des résolutions plus grandes. Cela permet une analyse plus complète de la texture, offrant ainsi une représentation plus robuste et précise des motifs dans l'image. L'étude de l'état de l'art montre que les variantes LBP (*Local Binary Pattern*) constitue un choix populaire et pertinent.

- Méthodes basées sur des transformations fréquentielles : sont utilisées pour analyser et extraire des caractéristiques texturales à différents niveaux de détail. En particulier, celles qui décomposent l'image en sous-bandes fréquentielles et offrent plusieurs résolutions d'images, telles que les transformées en ondelettes, permettent une analyse de la texture à différentes échelles, capturant ainsi des caractéristiques riches et variées. En effet, l'analyse multi-échelle permet d'examiner la texture sous différentes granularités, offrant une meilleure compréhension des structures globales tout en mettant en évidence les détails fins et les structures orientées. Cette analyse est particulièrement utile dans des applications où les objets ou les motifs peuvent être observés sous divers angles, tailles ou orientations. Cependant, certaines de ces méthodes peuvent engendrer un surplus de données et accroître la complexité d'implémentation. À cet égard, la transformée en ondelettes stationnaire et bi-dimensionnelle SWT (*Stationnary Wavelet Transform*) constitue une variante intéressante, offrant une décomposition multi-échelles et directionnelle. La SWT est largement utilisée en traitement d'images en raison de sa capacité à préserver les dimensions spatiales de l'image à chaque niveau de décomposition, contrairement à la transformée en ondelettes discrètes (DWT) classique qui applique un sous-échantillonnage à chaque décomposition [42], [43]. Cette propriété permet d'augmenter aisément le nombre de niveaux de résolution afin de réaliser une analyse approfondie sans se heurter aux inconvénients liés à la réduction progressive des dimensions des sous-bandes obtenues, ou aux artefacts provoqués par le filtrage aux bords. En effet, avec la SWT, toutes les sous-bandes générées possèdent la même taille que l'image originale.

- Méthodes basées sur le *transfer learning* : exploitent l'apprentissage profond via la puissance des réseaux de neurones convolutifs pré-entraînés pour analyser et extraire des caractéristiques texturales à partir d'images. En particulier, elles utilisent des modèles d'apprentissage profond construits sur d'importants ensembles de données génériques pour capturer de manière hiérarchique des informations à différents niveaux de détail, depuis les caractéristiques de bas niveau (bords, textures locales) jusqu'aux représentations de haut niveau (structures globales et contextuelles). La capacité du *transfer learning* à générer des représentations discriminantes en fait une alternative, même si elle demeure généralement moins interprétable que les méthodes *hand-crafted*.

Les sections qui suivent présentent une description des méthodes d'extraction de descripteurs sélectionnées dans le but d'atteindre les objectifs énoncés pour cette étude.

3.3. Les descripteurs LBP

Les descripteurs LBP (*Local Binary Patterns*) constituent une méthode d'extraction de caractéristiques de texture qui repose sur la capture des motifs locaux dans une image en niveaux de gris. Ils sont largement utilisés en raison de leur simplicité, de leur efficacité et de leur robustesse face aux variations d'éclairage et d'orientation.

Le descripteur LBP est basé sur la comparaison entre le pixel central d'un voisinage et ses pixels voisins. Pour chaque pixel central, les niveaux de gris des pixels voisins sont comparés à celui du pixel central et, selon la comparaison (supérieure ou inférieure), un motif binaire est créé. Ce motif binaire est ensuite converti en une valeur décimale unique appelée code LBP. Un histogramme de ces codes LBP est alors généré pour caractériser la distribution des motifs de texture dans l'image [20]. Dans le cas où le voisinage est sur 8 pixels (voir la figure 5) le motif binaire aura 8 bits et l'histogramme résultant aura 256 bins (2^8 =256).

Les principaux paramètres influençant les descripteurs LBP sont :

- **P** : nombre de voisins autour du pixel central. Ce paramètre affecte la quantification de l'espace angulaire, permettant de capturer des détails directionnels ;
- **R** : rayon du voisinage autour du pixel central. Ce paramètre détermine l'échelle spatiale des motifs capturés. Des valeurs plus élevées de R permettent de capturer des motifs de texture plus larges, ce qui est essentiel pour une analyse multi-échelle.

Invariance à la rotation et motifs uniformes *:* l'invariance à la rotation, obtenue par décalage circulaire des motifs binaires, permet d'identifier des textures indépendamment de leur orientation. Les motifs uniformes, qui comportent aux plus deux transitions entre 0 et 1, réduisent la taille du descripteur tout en conservant les informations essentielles. La fonction MATLAB « extractLBPFeatures », utilisée dans notre étude, intègre ces deux concepts afin d'optimiser l'extraction des caractéristiques [2].

Taille du descripteur LBP : la taille du vecteur descripteur dépend du choix d'utiliser ou non les motifs uniformes et/ou l'invariance à la rotation. Voici une analyse détaillée pour P = 8 voisins en se basant sur les définitions précises et les résultats connus dans la littérature.

- LBP standard (sans motifs uniformes et sans invariance à la rotation) : Dans le LBP standard, chaque voisin peut prendre deux valeurs (0 ou 1), générant 2^P combinaisons binaires possibles. La taille de l'histogramme résultant est alors $2^P = 256$. L'histogramme contient donc 256 bins, chacun correspondant à un motif unique. Bien que cet histogramme capture toute la variabilité des motifs locaux, il est volumineux et inadapté pour des applications nécessitant une faible dimensionnalité.

- LBP avec motifs uniformes (sans invariance à la rotation) : Les motifs uniformes sont définis comme ceux ayant, aux plus, deux transitions entre 0 et 1 dans une représentation circulaire. Pour P=8, il existe exactement 58 motifs binaires uniformes distincts sans invariance à la rotation. Tous les motifs non uniformes (plus de 2 transitions) sont regroupés dans un bin unique. Ainsi, la taille totale de l'histogramme résultant est 59 bins (58 motifs uniformes distincts et un 1 motif non uniforme), préservant les informations de texture les plus pertinentes.

- LBP avec motifs uniformes et invariance à la rotation : Lorsque l'invariance à la rotation est prise en compte, les motifs équivalents sous rotation circulaire sont regroupés, ce qui réduit l'histogramme résultant à 10 bins et offre une robustesse accrue aux variations d'orientation dans les textures. Ainsi, parmi les 58 motifs uniformes, les motifs équivalents sont réduits à 9 motifs rotationnellement invariants : un seul motif sans transition (00000000) et 8 motifs avec deux transitions. Par ailleurs, les motifs non uniformes sont toujours regroupés dans un bin unique.

Finalement, pour un voisinage P=8, la taille du descripteur passe de 256 bins dans le LBP standard à 59 bins avec les motifs uniformes, et à 10 bins en ajoutant l'invariance à la rotation. Ces ajustements rendent le descripteur LBP plus compact et efficace, particulièrement adapté pour des applications de recherche et de reconnaissance de texture nécessitant une faible dimensionnalité et une robustesse aux transformations géométriques [20].

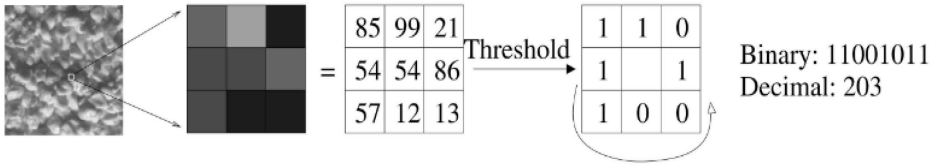
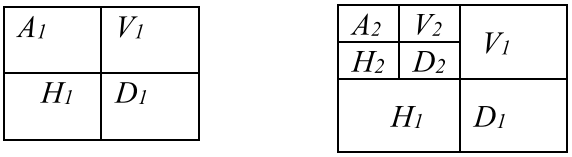


Figure 4: Illustration du processus de conversion des valeurs de pixels en un motif binaire dans l'algorithme LBP [20].

3.4. Les descripteurs SWT

Les descripteurs SWT reposent sur la décomposition d'une image en sous-bandes fréquentielles par la transformée en ondelettes stationnaires (SWT), puis sur l'extraction de caractéristiques à partir des sous-bandes obtenues (telles que l'énergie, l'écart-type, la moyenne, l'entropie, etc.). Cette approche multi-résolution est particulièrement utile pour analyser les textures à différentes échelles, car elle permet d'examiner à la fois les structures globales et les détails fins de l'image. La SWT est largement utilisée en traitement d'images en raison de sa capacité à préserver les dimensions spatiales de l'image à chaque niveau de décomposition, contrairement à la transformée en ondelettes discrètes (DWT) classique qui applique un sous-échantillonnage [42], [43].

La SWT décompose une image prétraitée en plusieurs sous-bandes en appliquant de manière itérative un banc de filtres passe-bas et passe-haut, sans effectuer de réduction de taille. Ainsi, à chaque niveau de décomposition l, l'image est divisée en quatre sous-bandes, H_l, V_l, D_l et A_l toutes de même taille que l'image d'origine, mais avec un contenu fréquentiel distinct et localisé, comme l'illustre la figure 5 ci-dessous :



(a) (b)

Figure 5** : **Partition fréquentielle en sous-bandes dans une décomposition d'une image à l'aide de la SWT : a. à un niveau, b. à deux niveaux.

- A_l (Approximation) : Sous-bande de basses fréquences en horizontal et vertical, représentant une version lissée de l'image.
- H_l (Détails horizontaux) : Capte les hautes fréquences verticales de l'image et met en évidence ses structures dans la direction horizontale.
- V_l (Détails verticaux) : Capte les hautes fréquences horizontales de l'image et met en évidence ses structures dans la direction verticale.
- D_l (Détails diagonaux) : Capte les hautes fréquences diagonales et met en évidence les structures diagonales de l'image.

Ces sous-bandes conservent les mêmes dimensions spatiales que l'image d'entrée $M \times N$ (par exemple, 576 × 576), ce qui permet une extraction cohérente et précise des caractéristiques texturales à chaque niveau de résolution. Les sous-bandes $H_l(i,j)$, $V_l(i,j)$ et $D_l(i,j)$, correspondant à un niveau de décomposition l, sont fusionnées en calculant l'énergie locale combinée selon la formule suivante :

$$E_l(i,j) = \sqrt{H_l(i,j)^2 + V_l(i,j)^2 + D_l(i,j)^2}$$

L'agrégation des détails par le calcul de l'énergie permet de capturer une signature isotrope, atténuant l'effet des rotations et des orientations des motifs de texture dans l'image.

À partir de cette énergie, deux statistiques globales sont calculées pour chaque niveau de résolution l :

– L'écart-type, qui mesure la variabilité des énergies locales

$$\mu_l = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N} E_l(i,j)$$

– La moyenne**,** qui quantifie la puissance globale de la texture à chaque niveau de résolution *l*

$$\sigma_l = \sqrt{\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}(E_l(i,j) - \mu_l)^2}$$

Le même calcul de moyenne et d'écart-type s'applique également à la sous-bande d'approximation A_l.

Ces caractéristiques statistiques sont ensuite concaténées sur l'ensemble des niveaux de résolution de la SWT afin de former les composantes du vecteur descripteur multi-échelle comme l'indique l'expression ci-après :

$$f_{SWT} = [\,\mu A1,\ \sigma A1,\ \mu 1,\ \sigma 1,\ \mu 2,\ \sigma 2, \dots,\ \mu L, \sigma L\,].$$

La performance de la SWT dépend principalement de deux paramètres :

– Nombre de niveaux de décomposition *L* : Ce paramètre détermine la profondeur de l'analyse multi-échelles ou multi-résolutions de la texture. Un nombre plus élevé permet de capturer les textures à différentes échelles, des détails les plus fins aux structures plus grossières.

– Type de filtre d'ondelette : L'ondelette Haar est souvent privilégiée pour sa simplicité et sa rapidité d'exécution, mais d'autres filtres, comme ceux de Daubechies, peuvent mieux modéliser des variations plus complexes dans l'image [13], [43].

La taille du vecteur descripteur SWT dépend directement du nombre de niveaux de décomposition *L* et du nombre de caractéristiques extraites. À chaque niveau de décomposition, deux mesures sont calculées (écart-type et moyenne de l'énergie) à partir des sous-bandes de détails, auxquelles s'ajoutent deux caractéristiques d'énergie issues de la sous-bande d'approximation retenue. La taille totale du descripteur est donc donnée par la formule $2 + 2 \times L$. Par exemple, pour une décomposition à 4 niveaux, le vecteur descripteur reste compact, avec seulement 10 éléments.

3.5. Les descripteurs par transfer learning

Ces descripteurs s'appuient sur l'utilisation des réseaux de neurones convolutionnels VGG16 et ResNet50, pré-entraînés sur les jeux de données visuelles d'ImageNet (ILSVRC-2012), qui

comprend plus d'un million d'images réparties en 1000 catégories d'objets [9]. Le choix de ces modèles est motivé à la fois par leur disponibilité dans l'environnement de simulation MATLAB et par l'accessibilité des ressources matérielles nécessaires à leur exécution. Chacun de ces réseaux est utilisé pour extraire automatiquement, à partir d'une image soumise à l'analyse, un vecteur descripteur nommé respectivement VGG16 ou ResNet50, selon le modèle mobilisé. Contrairement aux méthodes *hand-crafted*, qui reposent sur des algorithmes explicites tels que LBP ou SWT, cette méthode exploite des modèles ayant déjà appris, sur un large corpus d'images, à capturer des caractéristiques visuelles discriminantes à différents niveaux de détail.

La procédure d'extraction commence par un prétraitement de l'image d'entrée afin de l'adapter aux contraintes du réseau utilisé. Cette étape comprend :

- Le redimensionnement de l'image aux dimensions attendues (224 × 224 pixels pour VGG16 et ResNet50) ;
- La normalisation de l'intensité, et, le cas échéant, la conversion d'images en niveaux de gris en images pseudo-RGB par duplication des canaux.

Afin de renforcer l'invariance aux rotations, une stratégie d'augmentation de données est appliquée. L'image d'origine est soumise à plusieurs rotations (par exemple, 0°, 90°, 180° et 270°) produisant plusieurs versions orientées différemment. Pour chacune de ces versions, le réseau pré-entraîné extrait un vecteur de caractéristiques (ou des activations) à partir d'une couche intermédiaire sélectionnée, généralement une couche riche en information sémantique. Les vecteurs ainsi obtenus sont ensuite agrégés, par calcul de la moyenne, afin de produire le vecteur descripteur final, plus robuste aux variations d'orientation de l'image. La figure 6 illustre ce processus.

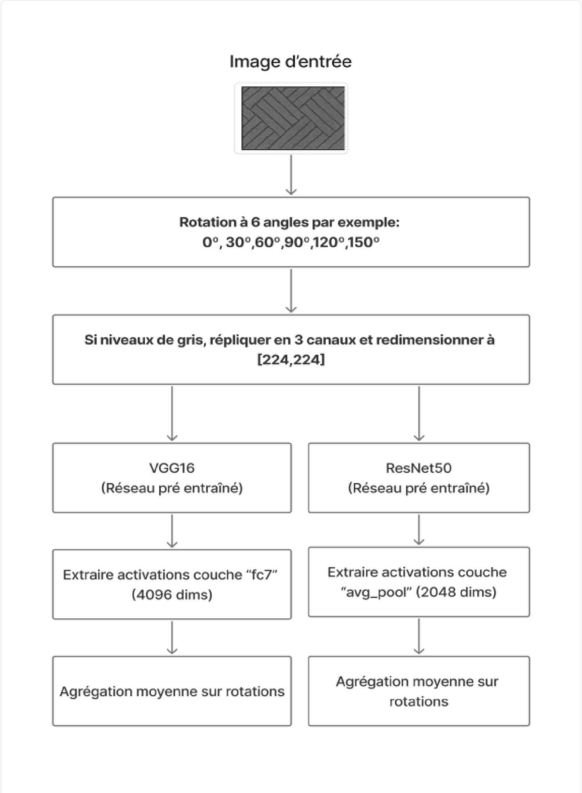


Figure 6 : Pipeline d'extraction de descripteurs profonds avec VGG16 et ResNet50.

Plusieurs paramètres influencent la qualité des descripteurs VGG16 et ResNet50 obtenus :

- Le choix du réseau pré-entraîné (par exemple, VGG16 ou ResNet50) ainsi que celui de la couche d'extraction (*fc7* pour VGG16 ou *avg_pool* pour ResNet50), qui déterminent à la fois la richesse sémantique et la dimension du vecteur descripteur ;
- La stratégie d'augmentation par rotations, incluant le nombre d'angles considérés et la méthode d'agrégation des vecteurs extraits (par exemple, la moyenne) ;
- La qualité du prétraitement (redimensionnement, normalisation), qui garantit la cohérence des entrées soumises au réseau.

Les descripteurs VGG16 et ResNet50 obtenus par apprentissage par transfert offrent la richesse des représentations hiérarchiques acquises lors de l'entraînement sur de larges ensembles de données. Cette approche permet de capturer à la fois des informations locales et globales, contribuant ainsi à une meilleure robustesse aux variations d'orientation.

3.6. Approche hybride

Dans le contexte de notre projet intitulé « descripteurs de textures d'images avec invariance à la rotation et approche multi-échelles pour la recherche d'images par le contenu », nous proposons

de tirer parti de la complémentarité entre plusieurs méthodes d'extraction de caractéristiques afin de maximiser les performances de recherche et de récupération d'images.

Les descripteurs LBP présentent une robustesse naturelle face aux rotations locales, en particulier grâce à leurs variantes invariantes à la rotation tel que le RI-LBP. De leur côté, les descripteurs SWT permettent une analyse multi-échelle efficace tout en conservant la structure spatiale de l'image à chaque niveau de décomposition. Contrairement à la DWT classique, la SWT ne procède à aucun sous-échantillonnage, ce qui garantit une stabilité spatiale favorable à une agrégation cohérente des énergies issues des sous-bandes de détails.

Dans cette optique, nous formulons l'hypothèse que la combinaison des descripteurs LBP et SWT, via la concaténation pondérée de leurs vecteurs descripteurs**,** permettrait de bénéficier simultanément :

- D'une invariance renforcée à la rotation, apportée par les motifs locaux du RI-LBP,
- Et d'une description multi-échelle des textures**,** rendue possible par l'extraction de mesures d'énergies à différents niveaux de résolution dans la décomposition SWT comme l'exprime la formule ci-dessous :

$$f\_hybride = [w0 * f\_LBP, (1 - w0) * f\_SWT]$$

Afin d'uniformiser les échelles des deux vecteurs avant mesure de similarité, une normalisation statistique de type *z-score* est appliquée au vecteur fusionné, après la combinaison pondérée. Cette normalisation consiste à centrer chaque dimension du descripteur en soustrayant sa moyenne, puis à la réduire en divisant par son écart type. Elle permet d'éviter qu'une composante domine les autres du fait de l'amplitude de ses valeurs numériques, et contribue à améliorer la stabilité et la comparabilité des distances entre descripteurs. Elle constitue donc un paramètre important dans l'efficacité globale du système CBIR, dont l'influence est prise en compte lors de l'analyse expérimentale.

Cette approche vise ainsi à construire un vecteur descripteur hybride, discriminant, compact, robuste et adapté à la recherche d'images texturées présentant des variations d'orientation et de structure.

En parallèle, les descripteurs issus du *transfer learning* constituent une alternative moderne et puissante aux méthodes traditionnelles. Il offre notamment une capacité d'abstraction élevée, bien que dépendant davantage des ressources matérielles et de la taille des données d'apprentissage.

Enfin, une analyse comparative est envisagée entre :

- La méthode hybride *hand-crafted* fondée sur la combinaison LBP + SWT,
- et la méthode par *transfer learning* (à l'aide des modèles VGG16 et ResNet50)

En évaluant leur précision, leur robustesse aux variations géométriques, ainsi que leur complexité computationnelle. Cette analyse comparative permettra de dégager les forces et les limites de chaque méthode.

3.7. Les distances pour le classement de similarité

Les distances jouent un rôle essentiel dans le processus de classement des images en fonction de leur similarité, particulièrement dans les systèmes CBIR. Deux distances principales sont proposées pour ce projet. Cette section décrit leur rôle, leur pertinence vis-à-vis des descripteurs utilisés.

Distance Euclidienne : cette métrique est couramment utilisée pour évaluer la similarité entre deux vecteurs de caractéristiques. Avec les descripteurs LBP, la distance euclidienne permet de comparer directement les histogrammes des codes LBP en tenant compte de la magnitude des différences. Pour les descripteurs SWT, elle capture efficacement les variations entre les mesures de caractéristiques issues des sous-bandes de décomposition, offrant une mesure simple et rapide pour l'analyse de similarité.

Mathématiquement, pour deux vecteurs $A = [a_1, a_2, ..., a_n]$ et $B = [b_1, b_2, ..., b_n]$, elle est exprimée par :

$$d_{Eucl}(A,B) = \sqrt{\sum_{i=1}^{n}(a_i - b_i)^2}$$

Distance du Khi-Carré (Chi-Square) : est une mesure statistique souvent utilisée pour comparer deux histogrammes ou distributions. Elle est définie comme suit pour deux vecteurs A et B :

$$d_{x^2}(A,B) = \sum_{i=1}^{n}\frac{(a_i - b_i)^2}{a_i + b_i + \in}$$

où ϵ est un terme de régularisation pour éviter la division par zéro. Cette distance est particulièrement utile pour des données d'histogrammes, car elle pondère les différences en fonction des valeurs des éléments, mettant davantage l'accent sur les faibles différences. Avec les descripteurs LBP, cette métrique est bien adaptée pour comparer les histogrammes, car elle capture les relations proportionnelles entre les bins. Pour les descripteurs SWT, son utilisation est moins directe, car les caractéristiques mesurées dans les sous-bandes ne représentent pas des fréquences.

3.8. Bases de données pour l'évaluation des systèmes CBIR

Les performances d'un système CBIR pour la recherche d'images de textures sont évaluées sur diverses bases de données de référence :

- Kylberg Texture Dataset : Utilisé pour tester des descripteurs de texture invariants à la rotation [12], [36] ;
- VisTex et Brodatz : Bases de données standard pour les tests de descripteurs de texture [11], [34] ;
- Outex est une base de données largement utilisée dans le domaine de l'analyse de textures pour évaluer les performances des algorithmes de classification et de reconnaissance et est principalement conçue pour tester la robustesse des méthodes face à des variations comme l'éclairage, la rotation ou le bruit [22], [35].

Le choix des bases suivantes découle de leur complémentarité et de leur adéquation avec les objectifs du présent projet, notamment l'analyse de textures invariantes à la rotation et l'étude de la robustesse multi-échelle des descripteurs.

Les performances de recherche et récupération du CBIR seront mesurées en termes de précision et de rappel (voir la *section 2.6* de l'état de l'art). Ces mesures seront appliquées pour les bases de données de référence utilisées.

4. DEVELOPPEMENT ET RÉSULTATS

L'objectif de ce chapitre est de valider le bon fonctionnement du système CBIR et des méthodes d'extraction de caractéristiques de textures, à savoir les méthodes *hand-crafted* (LBP, SWT et hybride) et les méthodes basées sur le *transfer learning* (VGG16 et ResNet50). Des expérimentations sont effectuées pour tester l'efficacité du système CBIR sur diverses bases de données de textures.

4.1. Bases de données et environnement de développement

4.1.1. VisTex

La base de données VisTex [34] est une collection d'images en couleur conçue pour tester les performances des descripteurs de textures sur une large gamme de motifs texturaux. Elle contient un total de 640 images en format RGB, réparties en 40 classes distinctes, avec 16 images par classe. Chaque image a une résolution spatiale de 128×128 pixels. Les textures présentes dans cette base de données couvrent une diversité de motifs naturels et artificiels, ce qui permet de valider les capacités des descripteurs à capturer les caractéristiques et à exploiter les informations chromatiques pour améliorer les résultats.

Figure 7 : Exemples d'images de texture de la base de données VisTex [23].

4.1.2. Outex

La base de données Outex_TC_00010 [35] est spécifiquement conçue pour évaluer la robustesse des algorithmes de classification de textures face aux variations de rotation. Elle se compose de 24

classes de textures, chacune capturée sous neuf angles de rotation différents : 0°, 5°, 10°, 15°, 30°, 45°, 60°, 75° et 90°. Pour chaque classe et chaque angle, 20 images sont disponibles, totalisant ainsi 4 320 images (24 classes × 9 angles × 20 images). Les images sont en niveaux de gris et ont une résolution spatiale de 128×128 pixels. Cet ensemble de données permet de tester l'invariance à la rotation des descripteurs de texture.

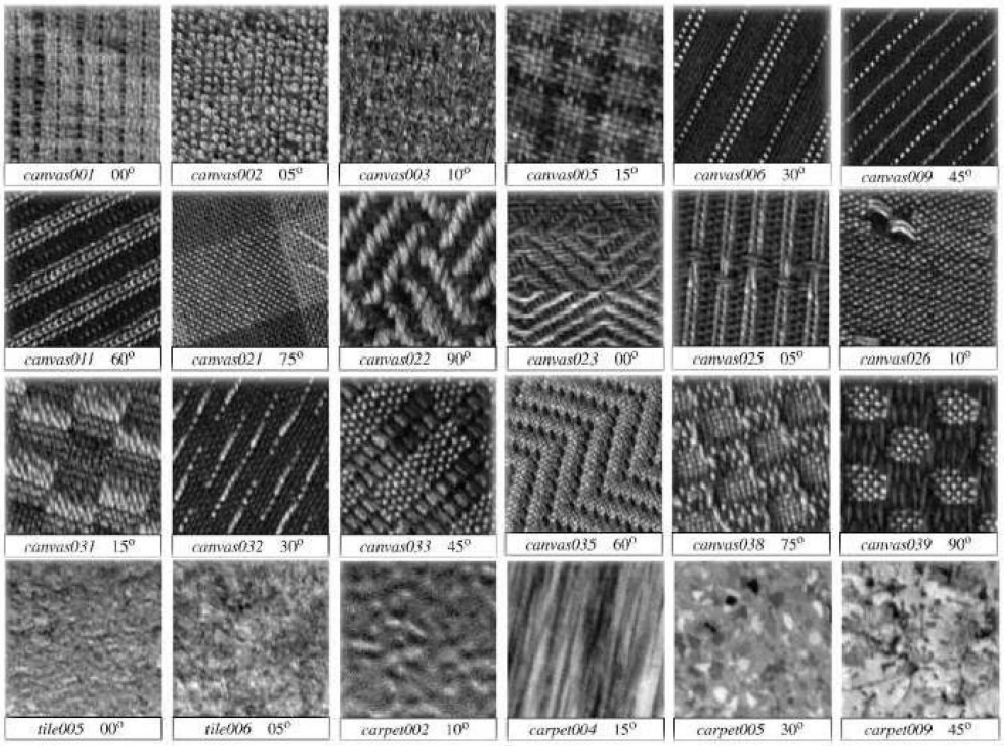


Figure 8 : Exemples d'images de texture de la base de données Outex [24].

4.1.3. Kylberg

La base de données Kylberg [36] est spécialement conçue pour évaluer la robustesse des algorithmes de classification de textures face aux variations de rotation. Elle se compose de 28 classes de textures naturelles, chacune capturée sous 12 angles de rotation différents (par exemple : 0°, 30°, 60°, 90°, 120°, 150°, …). Pour chaque classe et pour chaque angle, 160 images sont disponibles, totalisant ainsi 53 760 images (28 classes × 12 angles × 160 images). Les images, en niveaux de gris, présentent une résolution spatiale de 576×576 pixels. Cet ensemble de données offre un cadre exhaustif pour tester l'invariance à la rotation des descripteurs de texture.

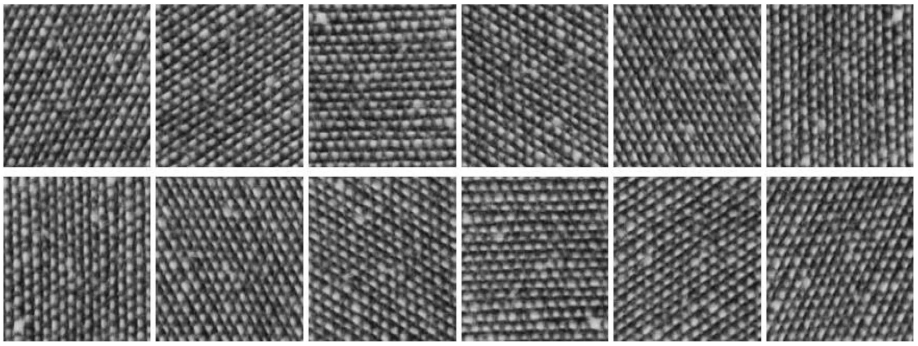


Figure 9 : Exemple d'un échantillon de texture de la classe « cushion » et de ses douze versions tournées (de 0° à 330°) [12].

4.1.4. Configuration matérielle et logicielle

Pour la mise en œuvre d'un CBIR permettant différentes expérimentations, la configuration matérielle et logicielle se compose des éléments suivants :

- **Logiciels :** MATLAB ;
- **Matériel :** Processeur Intel i7, 8 Go de RAM ;
- **Bibliothèques :** Utilisation de diverses fonctions MATLAB issues des *toolbox* telles qu'*Image processing*, *Computer vision* et *Deep learning*.

4.2. Protocole expérimental

L'évaluation des méthodes proposées a été effectuée sur trois bases de données de référence en analyse de textures : VisTex, Outex et Kylberg. Chacune de ces bases présente des défis spécifiques liés à la diversité des textures, aux variations d'orientations et au volume d'images. Leur structure permet également de disposer d'une vérité terrain explicite, essentielle pour la conduite des tests et l'évaluation des performances. Les caractéristiques principales de ces bases sont les suivantes :

- VisTex : 40 classes (sans variation d'orientation) × 16 images = 640 images
- Outex : 24 classes × 9 rotations × 20 images = 4 320 images
- Kylberg : 28 classes × 12 rotations × 160 images = 53 760 images

Un prototype de système CBIR est développé selon une architecture conventionnelle à deux phases (voir la figure 1) : une phase hors ligne et une phase en ligne. Ce prototype intègre la méthode d'extraction de descripteurs sélectionnée et utilise une métrique de distance comme mesure de similarité. Les méthodes d'extraction de descripteurs mises en œuvre sont réparties en deux grandes catégories :

- Méthodes *hand-crafted* :
  - SWT (Ondelettes stationnaires, analyse multi-échelle)
  - LBP (Motif binaire local, invariant à la rotation)
  - Hybride LBP+SWT (Fusion linéaire pondérée des 2 descripteurs, poids w_0 optimisé)
- Méthodes par *transfer learning*:
  - VGG16 (réseau de neurones convolutionnel pré-entraîné)
  - VGG16 + agg (réseau de neurones convolutionnel pré-entraîné, avec agrégation multi-angle)
  - ResNet50 (réseau de neurones convolutionnel pré-entraîné)
  - ResNet50+agg (réseau de neurones convolutionnel pré-entraîné, avec agrégation multi-angle)

    Pour VGG16 et ResNet50, l'agrégation multi-angle consiste à calculer la moyenne des descripteurs extraits à partir de 15 orientations différentes de l'image : [0°, 15°, 30°, 45°, 60°, 90°, 120°, 150°, 180°, –150°, –120°, –90°, –60°, –45° et –30°], dans le but de renforcer l'invariance à la rotation.

Pour évaluer l'efficacité d'une expérience de recherche CBIR, chaque image de la base de données de test est utilisée comme requête dans le système CBIR. Ce protocole garantit une couverture complète et une profondeur de recherche proportionnelle au volume de données à explorer. Comme la vérité terrain (*ground truth*) est connue pour chaque requête, l'efficacité de la recherche pour chaque requête est mesurée à l'aide de deux indicateurs standards : le rappel (*recall*) et la précision (*precision*).

Le rappel correspond au pourcentage d'images pertinentes retrouvées parmi les N premières images retournées (TopN) par le CBIR. Une image est considérée comme pertinente si elle appartient à la même classe que l'image requête. Tous les résultats de recherche d'images présentés dans ce travail sont obtenus en moyennant les valeurs de rappel et de précision sur l'ensemble des requêtes effectuées (précision globale et rappel global). Par exemple, une expérience de recherche d'images sur la base de données Kylberg requiert l'exécution de 53 760 requêtes.

En complément, le temps moyen d'exécution par image est estimé et des valeurs moyennes de rappel et de précision par classe sont également calculées, afin de permettre une analyse plus fine des performances selon les classes de textures. À noter que lorsque la valeur du TopN est égale à la taille des classes de la base de données test, les valeurs de rappel et de précision sont alors égales.

4.2.1 Comparaison des distances de similarité : euclidienne vs chi-carré

Des tests comparatifs ont été réalisés pour évaluer les performances des distances euclidienne et khi-carré sur les descripteurs LBP et SWT dans le cadre du classement de similarité des images. Les résultats ont montré que :

- La distance euclidienne offre une meilleure performance globale en termes de précision du classement pour les deux descripteurs, notamment en raison de sa simplicité et de sa capacité à capturer les différences globales entre les vecteurs de caractéristiques [2], [7] ;
- La distance du khi-carré, bien qu'efficace pour des histogrammes, n'a pas surpassé la distance euclidienne dans ce contexte. Elle a montré des performances légèrement inférieures, ce qui rejoint les conclusions de plusieurs travaux antérieurs soulignant sa sensibilité aux faibles valeurs dans les histogrammes [2], [21].

À la suite des résultats des tests comme le montre la figure 10 ci-dessous, la distance euclidienne a été retenue comme mesure principale pour le classement de similarité des images dans ce projet. Ce choix repose sur les raisons suivantes :

- Robustesse **:** La distance euclidienne a démontré une meilleure précision pour les descripteurs utilisés (LBP et SWT) ;
- Simplicité **:** Elle est moins complexe à calculer, ce qui garantit une implémentation plus rapide et efficace ;
- Universalité **:** Contrairement à la distance du khi-carré, qui est spécifique aux données d'histogrammes, la distance euclidienne est plus universelle et adaptée à différents types de descripteurs.

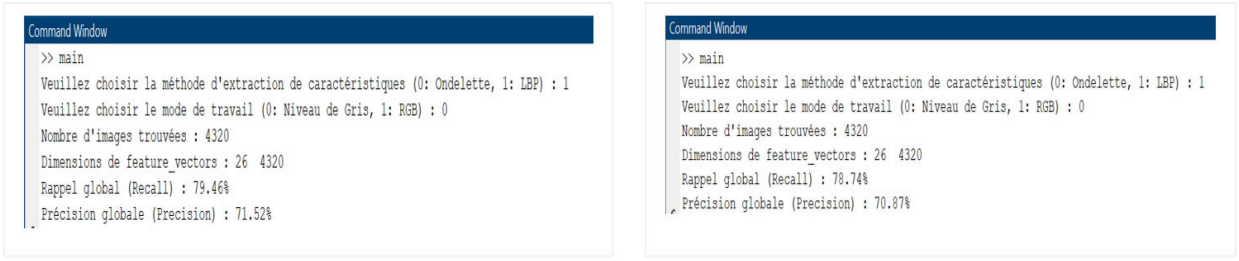


Figure 10 : Comparaison de l'impact de la distance utilisée sur les résultats a. Distance euclidienne b. Distance khi-carré.

4.3. Résultats et analyse sur VisTex

VisTex ne comportant pas de rotations, les expérimentations n'incluent pas d'agrégation multi-angle pour les méthodes VGG16 ou ResNet50.

Les résultats présentés dans le tableau 2 ci- dessous correspondent à des tests faits dans le contexte d'un TopN=20, LBP (R= [1,3], P=18) et SWT (nlevels=3).

Tableau 2 : Performances globales par méthode et format (a. Niveaux de gris b. RGB) sur la base VisTex (TopN = 20).

Méthode	Rappel (%)	Précision (%)
LBP	82,77	66,22
SWT	72,19	57,75
Hybride (w_0=0,6)	87,51	70,01
ResNet50	96,75	77,40
VGG16	92,52	74,02

a

Méthode	Rappel (%)	Précision (%)
LBP	86,78	69,42
SWT	73,39	58,71
Hybride (w_0=0,6)	89,85	71,88
ResNet50	RGB	97,47
VGG16	93,40	74,72

b

Analyse comparative : Les méthodes de *transfer learning* sans agrégation (ResNet50 et VGG16) se démarquent nettement, atteignant jusqu'à 97,47 % de rappel en format RGB. Le passage du niveau de gris au RGB apporte un gain de 3 à 4 %, ce qui souligne l'importance de l'information couleur dans l'analyse de textures. Cette validation initiale sur VisTex confirme la pertinence des descripteurs proposés, avant d'aborder les défis liés à l'invariance à la rotation sur les bases Outex et Kylberg.

4.4. Résultats et analyse sur Outex

Le Tableau 3 ci-dessous résume les performances globales (rappel, précision, temps de calcul) de chaque méthode sur Outex.

Tableau 3 : Performances globales sur Outex (TopN=180).

Méthode	Rappel Global (%)	Précision Globale (%)	Temps moyen (s/image)	Type
SWT (nlevels=4)	75,30	75,30	0,0058	Hand-crafted
LBP (R= [4,8], P=24)	81,71	81,71	0,0093	Hand-crafted
Hybride LBP+SWT (w_0=0,8)	88,29	88,29	0,0171	Hand-crafted (fusion)
VGG16	71,85	71,85	0,443	Deep learning (TL)
VGG16+agg	88,26	88,26	4,301	Deep learning (TL+agg)
ResNet50	76,71	76,71	2,987	Deep learning (TL)
ResNet50+agg	91,38	91,38	4,769	Deep learning (TL+agg)

Analyse comparative : Dans la base Outex, chaque image de texture distincte est accompagnée de 8 versions orientées, correspondant aux angles de rotation suivants : 0°, 5°, 10°, 15°, 30°,

45°, 60°, 75° et 90°. Cette structure permet une évaluation du degré d'invariance à la rotation des descripteurs extraits.

- Parmi les méthodes *hand-crafted*, la fusion hybride (LBP+SWT) se distingue nettement, surpassant les performances des approches individuelles. Cela met en évidence la complémentarité des descripteurs LBP et SWT, ainsi que la pertinence de leur combinaison.
- Les méthodes de *transfer learning* sans agrégation multi-angle (VGG16, ResNet50) n'atteignent pas le niveau de performance de la meilleure méthode *hand-crafted*, ce qui traduit une sensibilité aux variations d'orientation.
- L'introduction d'une agrégation multi-angle permet de tirer pleinement parti des modèles de *transfer learning* tout en renforçant la robustesse des descripteurs face à la variabilité des orientations : la combinaison ResNet50+agg obtient la meilleure performance, avec un rappel et une précision de 91,38 %**,** surpassant toutes les autres méthodes, y compris la meilleure approche *hand-crafted*.
- Pour un taux de rappel fluctuant entre 88,29% (pour la méthode Hybride) et 91.38 (pour ResNet50+agg), il apparait clairement que les descripteurs utilisés démontrent une forte capacité de discrimination, même en présence d'une variabilité rotationnelle importante.

4.4.1 Analyse détaillée par classe de texture

Dans le tableau 4 ci-après, les valeurs représentent les pourcentages (%) de rappel ou de précision par classe de texture. Il convient de rappeler que la base de données Outex comprend 20 classes de textures distinctes. Chacune composée de 180 images, réparties selon 9 orientations différentes.

Tableau 4 : Rappel (%) par classe et par méthode (Outex, TopN=180).

Classe	SWT	LBP	Hybride	VGG16+agg	ResNet50+agg
1	56,36	91,70	90,40	86,36	77,28
2	89,89	88,87	95,17	98,72	99,90
3	92,61	94,04	98,59	91,39	93,57
4	83,97	82,88	97,67	84,75	88,83
5	99,99	90,83	100,00	81,47	98,64
6	99,82	99,77	100,00	82,39	99,14
7	88,59	87,93	97,77	82,73	78,70
8	99,54	99,58	100,00	99,64	99,81
9	97,62	93,01	99,50	99,99	100,00
10	28,73	69,70	64,70	88,89	84,11
11	50,44	96,91	98,55	95,37	89,11
12	95,18	94,22	99,29	97,13	94,94
13	86,44	73,63	89,27	83,51	89,69
14	79,08	59,93	83,65	82,08	77,91
15	61,92	43,24	53,00	71,14	76,06
16	36,07	42,76	45,59	55,54	73,68
17	65,27	76,20	90,06	72,27	83,31
18	70,01	71,47	80,57	87,03	96,56

19	60,89	72,06	80,38	96,65	97,68
20	51,65	75,10	78,02	97,78	96,51
21	88,70	90,06	95,53	99,92	99,95
22	62,87	93,40	94,02	98,70	99,51
23	96,11	90,41	96,03	92,07	99,38
24	65,45	83,33	91,28	92,72	98,87

L'analyse suivante met en évidence, pour chaque méthode d'extraction de descripteur, les cinq classes les mieux reconnues et les cinq les moins bien reconnues, sur la base des taux de rappel (%) :

SWT

- Top 5 : 5 (99,99%), 6 (99,82%), 8 (99,54%), 9 (97,62%), 12 (95,18%)
- Pire 5 : 10 (28,73%), 16 (36,07%), 11 (50,44%), 20 (51,65%), 22 (62,87%)

LBP

- Top 5 : 6 (99,77%), 8 (99,58%), 11 (96,91%), 12 (94,22%), 3 (94,04%)
- Pire 5 : 15 (43,24%), 16 (42,76%), 10 (69,70%), 14 (59,93%), 13 (73,63%)

Hybride LBP+SWT

- Top 5 : 5 (100,00%), 6 (100,00%), 8 (100,00%), 9 (99,50%), 12 (99,29%)
- Pire 5 : 16 (45,59%), 15 (53,00%), 10 (64,70%), 14 (83,65%), 13 (89,27%)

VGG16+agg

- Top 5 : 9 (99,99%), 21 (99,92%), 2 (98,72%), 8 (99,64%), 22 (98,70%)
- Pire 5 : 7 (82,73%), 6 (82,39%), 4 (84,75%), 14 (82,08%), 1 (86,36%)

ResNet50+agg

- Top 5 : 9 (100,00%), 21 (99,95%), 2 (99,90%), 8 (99,81%), 23 (99,38%)
- Pire 5 : 16 (73,68%), 15 (76,06%), 14 (77,91%), 7 (78,70%), 1 (77,28%)

Ainsi, les textures 8, 9 et 21 sont toujours bien reconnues, tandis que les textures 1, 10, 14, 15 et 16 demeurent systématiquement difficiles pour l'ensemble des méthodes ; la fusion hybride LBP+SWT améliore les performances des méthodes manuelles, mais seuls les réseaux profonds avec agrégation multi-angle (notamment ResNet50+agg) parviennent à la fois à maximiser le rappel global et à uniformiser les performances entre classes.

En effet, la classe 9 atteint 100 % avec ResNet50+agg, illustrant sa forte distinctivité, tandis que la classe 16 enregistre les performances les plus faibles (environ 45 %), reflet de sa complexité structurelle. La figure 11 ci-contre permet de visualiser clairement ces écarts d'un simple coup d'œil :

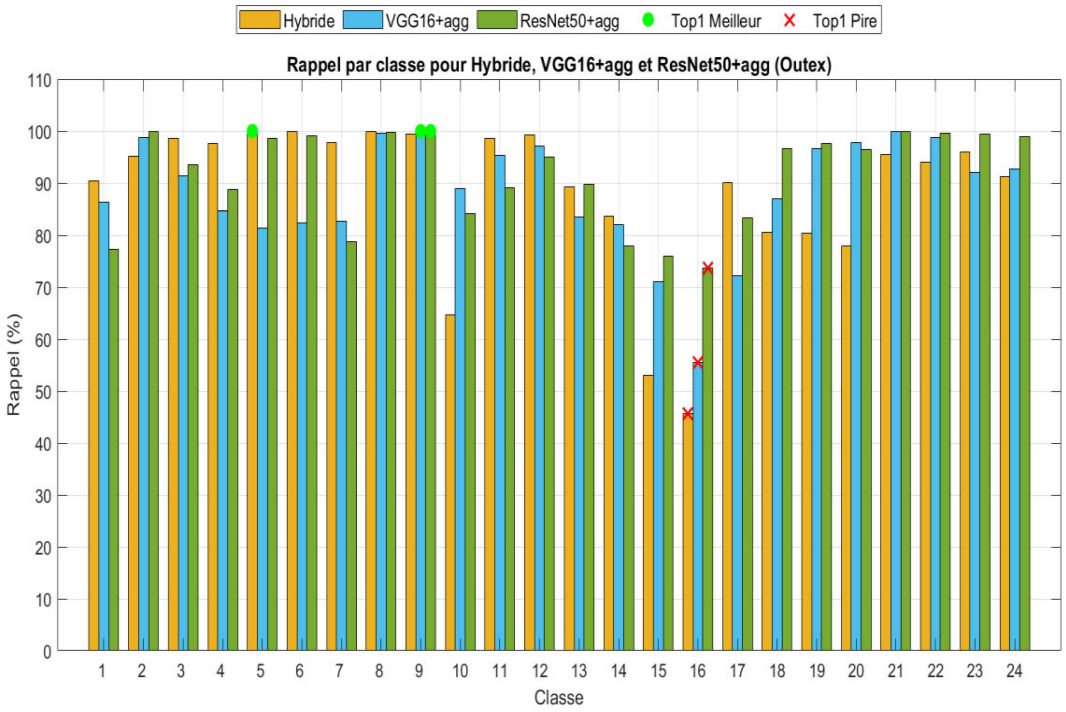


Figure 11 : Rappel par classe pour les méthodes Hybride, VGG16+agg et ResNet50+agg sur la base Outex

4.4.2. Impact du poids de fusion w_0 et effet de la normalisation

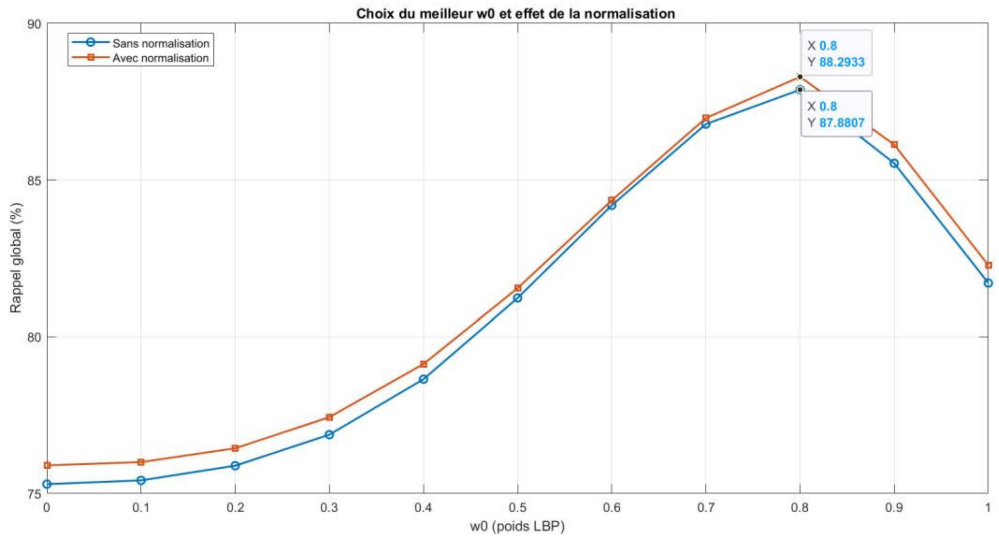


Figure 12 : Influence du poids de fusion w_0 sur la performance hybride, avec et sans normalisation (Outex).

Sur la base de données Outex, la fusion LBP+SWT atteint un pic de rappel à $w_0 = 0,8$ (≈ 88,3 %). La normalisation post-fusion rehausse le niveau de performance et atténue la sensibilité au choix de w_0. En pratique, $w_0 = 0,8$ avec normalisation constitue le meilleur compromis entre précision maximale et robustesse au calibrage.

4.4.3. Discussion globale et portée des résultats sur Outex

À la lumière des résultats obtenus sur la base Outex, il est pertinent de tirer les enseignements suivants concernant la performance comparative des méthodes évaluées :

- Les méthodes *hand-crafted* sont simples, très rapides et déjà robustes pour certaines classes de textures.
- La fusion hybride améliore significativement la performance globale des méthodes *hand-crafted*, et exhibe une capacité de discrimination élevée, même en présence de variabilité rotationnelle importante.
- Seules les méthodes de *transfer learning* avec agrégation multi-angle (surtout ResNet50+agg) permettent à la fois d'atteindre les meilleurs rappels, de réduire les écarts entre classes faciles et difficiles, et d'offrir une meilleure robustesse face à la variabilité rotationnelle.

Ainsi, le choix de la méthode doit s'appuyer sur un compromis entre la performance maximale (assurée par le *transfer learning* avec agrégation) et la simplicité / rapidité d'exécution (offerte par les méthodes *hand-crafted* hybrides), en fonction des contraintes et des objectifs de l'application visée**.**

4.5. Résultats et Analyse sur Kylberg

Le Tableau 5 ci-contre présente les performances globales (rappel, précision, temps de calcul) de chaque méthode sur la base Kylberg.

Tableau 5 : Performances globales sur Kylberg (TopN=1920).

Méthode	**Rappel Global (%)**	**Précision Globale (%)**	**Temps moyen (s/image)**	**Type**
SWT (nlevels=4)	75,40	75,40	0,159	Hand-crafted
LBP (R= [16,24], P=24)	74,80	74,80	0,117	Hand-crafted
Hybride LBP+SWT (w_0=0,9)	80,28	80,28	0,279	Hand-crafted (fusion)
VGG16	72,78	72,78	0,485	Deep learning (TL)
VGG16+agg	88,58	88,58	4,501	Deep learning (TL+agg)
ResNet50	78,52	78,52	3,109	Deep learning (TL)
ResNet50+ agg	92,09	92,09	4,808	Deep learning (TL+agg)

Analyse comparative : Dans la base Kylberg, chaque image de texture distincte est accompagnée de 12 versions orientées, correspondant aux angles de rotation suivants : 0°, 30°, 60°, 90°, 120°, 150°,180°, 210°, 240°, 270°, 300°, 330° et 360°. Cette configuration permet une évaluation du degré d'invariance à la rotation des descripteurs extraits. Compte tenu du volume de données par classe (1920 images), les constats suivants peuvent être formulés :

- La fusion hybride améliore significativement les résultats des méthodes *hand-crafted* les rendant supérieurs que ceux du *transfer learning* sans agrégation.
- Sur Kylberg, la distinction entre les méthodes *hand-crafted* et celles issues du *transfer learning* avec agrégation multi-angle est encore plus prononcée : seuls les réseaux profonds avec agrégation parviennent à dépasser les 90 % de rappel.

4.5.1. Analyse détaillée des rappels par classe

Dans le tableau 6 ci-contre, les valeurs représentent les pourcentages (%) de rappel ou de précision par classe de texture.

Tableau 6 : Rappel (%) par classe et par méthode (Kylberg, TopN=1920).

Classe	SWT	LBP	Hybride	VGG16+agg	ResNet50+agg
1	80,61	73,41	85,10	66,36	95,76
2	82,67	86,70	96,64	96,43	99,98
3	81,63	76,74	85,08	76,24	90,84
4	91,69	79,38	92,35	99,72	99,95
5	100,00	95,06	98,77	100,00	100,00
6	93,08	98,92	99,14	99,03	97,43
7	99,99	98,85	99,67	100,00	100,00
8	99,55	91,97	99,75	94,90	96,48
9	46,98	48,14	52,49	83,20	74,98
10	80,78	60,93	70,33	98,54	97,61
11	83,42	63,64	76,97	93,58	98,31
12	55,51	58,94	72,86	99,68	99,88
13	45,36	58,90	61,14	65,40	89,49
14	81,57	82,70	95,91	98,19	99,86
15	88,72	95,57	97,82	98,66	99,37
16	35,89	45,61	52,46	77,05	76,00
17	62,05	67,34	70,30	93,69	83,20
18	95,12	95,37	99,39	98,81	100,00
19	100,00	96,28	96,57	99,34	98,27
20	66,38	77,63	83,60	84,75	90,06
21	68,97	58,44	64,97	85,37	99,20
22	86,38	99,81	99,80	92,39	97,96
23	95,60	92,91	95,45	95,51	98,15
24	74,59	61,35	69,69	69,50	79,74
25	60,35	43,45	52,08	62,65	64,63
26	47,69	58,55	57,37	77,35	78,86
27	64,39	67,29	68,50	88,75	82,97
28	42,25	60,44	53,65	85,23	89,47

L’analyse met en évidence, pour chaque méthode d’extraction de descripteurs, les cinq classes les mieux reconnues et les cinq les moins bien reconnues, sur la base des taux de rappel (%) :

SWT

- Top 5 : 5 (100,00), 7 (99,99), 19 (100,00), 8 (99,55), 18 (95,12)
- Pire 5 : 28 (42,25), 13 (45,36), 16 (35,89), 26 (47,69), 9 (46,98)

LBP

- Top 5 : 6 (98,92), 7 (98,85), 22 (99,81), 15 (95,57), 5 (95,06)
- Pire 5 : 25 (43,45), 16 (45,61), 9 (48,14), 28 (60,44), 24 (61,35)

Hybride LBP+SWT

- Top 5 : 18 (99,39), 7 (99,67), 8 (99,75), 6 (99,14), 22 (99,80)
- Pire 5 : 25 (52,08), 28 (53,65), 13 (61,14), 16 (52,46), 9 (52,49)

VGG16+agg

- Top 5 : 5 (100,00), 7 (100,00), 12 (99,68), 4 (99,72), 15 (98,66)
- Pire 5 : 1 (66,36), 3 (76,24), 13 (65,40), 24 (69,50), 25 (62,65)

ResNet50+agg

- Top 5 : 5 (100,00), 7 (100,00), 2 (99,98), 18 (100,00), 4 (99,95)
- Pire 5 : 25 (64,63), 9 (74,98), 16 (76,00), 24 (79,74), 13 (89,49)

Les textures les plus distinctives (classes 5, 6, 7, 18, 22) sont systématiquement bien reconnues, tandis que les plus complexes (9, 13, 16, 24, 25, 28) demeurent problématiques, même pour les modèles de *Deep Learning* les plus performants.

La figure 13 ci-dessous permet de bien visualiser ces écarts en un coup d’œil :

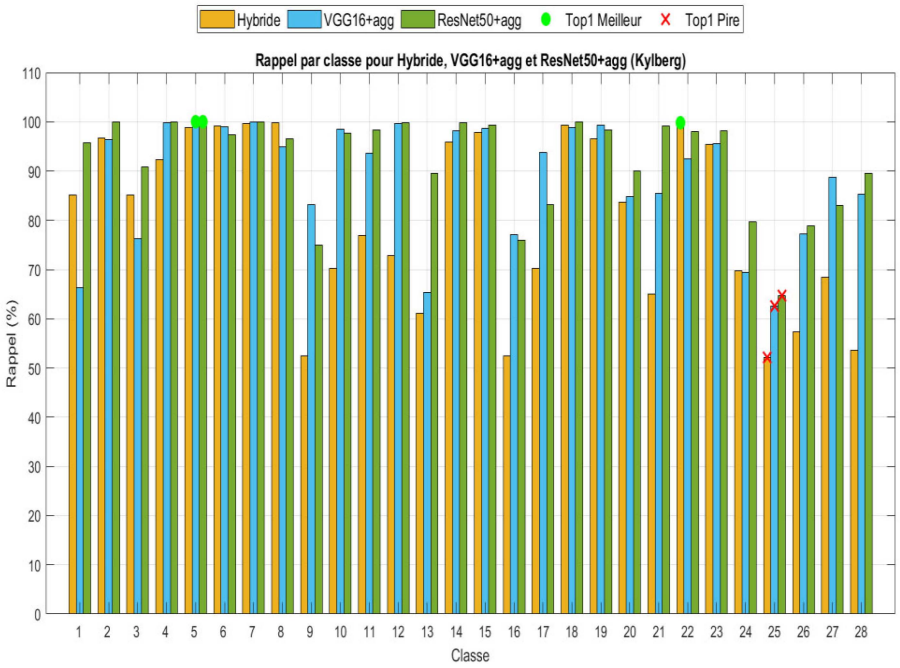


Figure 13 : Rappel par classe pour les méthodes Hybride, VGG16+agg et ResNet50+agg sur la base Kylberg

La classe 5 atteint près de 100 % de rappel quelle que soit la méthode, tandis que la classe 25 reste bloquée autour de 52 %.

4.5.2 Impact du poids de fusion w_0 et effet de la normalisation

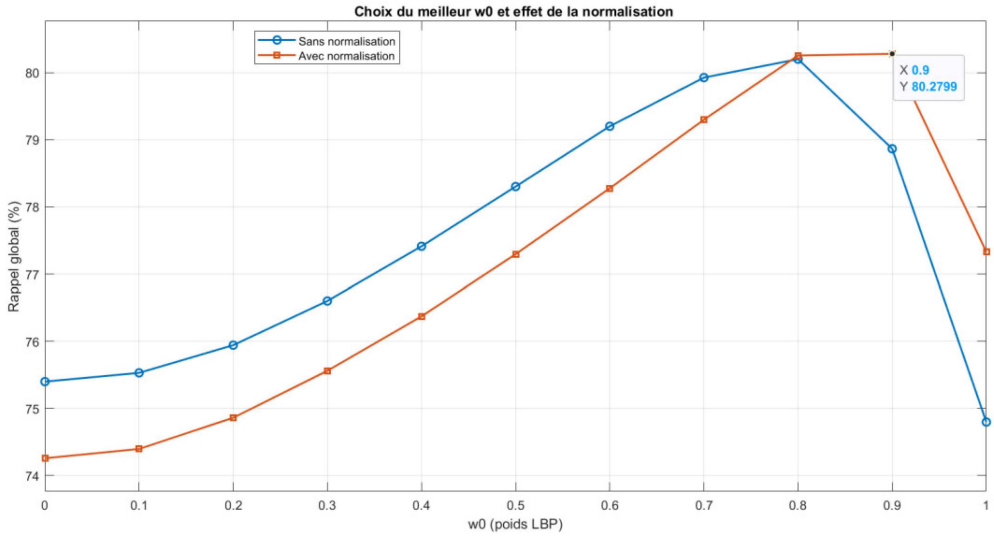


Figure 14 : Influence du poids de fusion w_0 sur la performance hybride, avec et sans normalisation (Kylberg)

Sur Kylberg, les rappels sans normalisation dominent pour $w_0 \leq 0,7$, mais s'inversent à partir de $w_0 = 0,8$ où la normalisation prend le dessus, avec un pic à $w_0 = 0,9$ (≈80,3 %). Ainsi, $w_0 = 0,9$ assorti de normalisation post-fusion offre le meilleur compromis entre performance maximale et stabilité au calibrage.

4.5.3. Discussion globale et portée des résultats sur Kylberg

La supériorité des méthodes de deep learning (transfer learning) avec agrégation multi-angle est manifeste sur Kylberg, une base volumineuse et complexe. Les hand-crafted/hybrides, même optimisées, montrent leurs limites dès que la diversité et le volume augmentent. L'analyse par classe confirme qu'il existe des textures structurellement complexes, quel que soit le descripteur.

4.6. Analyse exploratoire : effet de la transformation log-polaire sur SWT

Dans le but de renforcer l'invariance à la rotation, une transformation log-polaire a été appliquée en prétraitement aux images avant l'extraction des caractéristiques via la méthode SWT.

La figure 15 suivante montre l'effet de cette transformation sur une image de la base Kylberg. En remappant les coordonnées cartésiennes en log-polaire, les rotations deviennent des translations, ce qui est théoriquement favorable à l'invariance à la rotation.

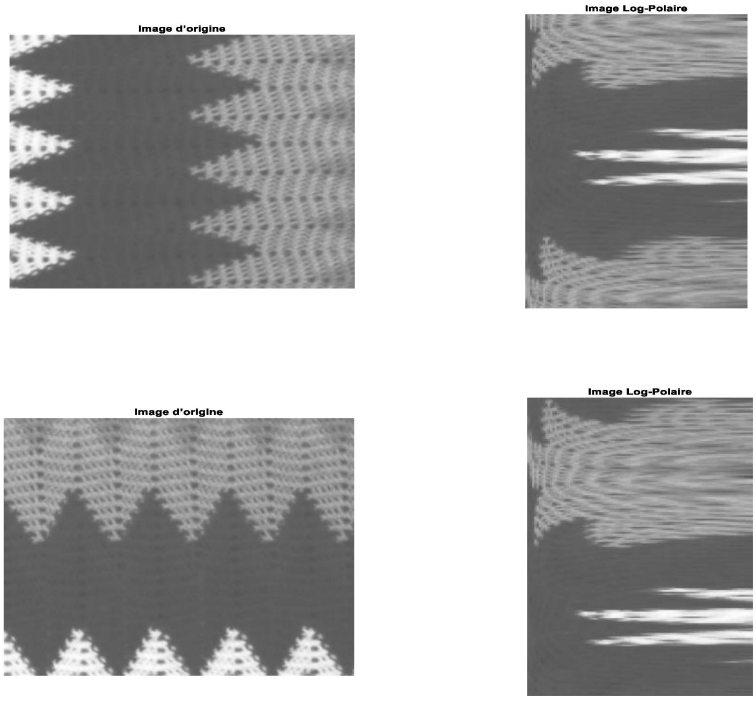


Figure 15 : Image blanket2-a-p001 et sa version rotée de 90° (gauche) et les images log-polaires correspondantes (droite).

Les performances de SWT avec et sans transformation log-polaire ont été comparées dans le tableau 7 ci-après :

Tableau 7 : Rappel global (%) du descripteur SWT avec et sans transformation log-polaire sur les bases Kylberg et Outex.

Base de données	Méthode	Rappel (%)
Kylberg	SWT	75,40
	SWT + log-polaire	65,43
Outex	SWT	75,30
	SWT + log-polaire	63,08

L'ajout de la transformation log-polaire a entraîné une baisse notable des performances en termes de rappel (%). Cette dégradation peut s'expliquer par des pertes d'information ou des distorsions liées au remappage. En l'état, cette approche n'a donc pas été retenue, bien qu'elle puisse rester prometteuse dans d'autres configurations, notamment avec des modèles adaptés à ce type de transformation.

4.7. Résumé des résultats

L'évaluation préliminaire sur la base VisTex, dépourvue de variations rotationnelles, a confirmé la supériorité des méthodes de *transfer learning*, justifiant ainsi le choix des descripteurs profonds pour cette étude. Les expérimentations menées sur Outex et Kylberg ont ensuite permis d'évaluer la capacité des différentes approches à gérer l'invariance à la rotation, axe central de notre travail.

Sur Outex, une base de taille moyenne, les descripteurs *hand-crafted*, en particulier la fusion LBP+SWT se montrent très compétitifs, tandis que les méthodes profondes ne déploient leur plein potentiel qu'en présence d'une agrégation multi-angle. À l'inverse, sur la base plus complexe et volumineuse de Kylberg, seules les approches à base de réseaux pré-entraînés avec agrégation parviennent à dépasser les limites des descripteurs issus des méthodes *hand-crafted*.

L'analyse par classe révèle que les textures les plus simples ou les plus ambiguës à représenter sont systématiquement les mêmes, quel que soit le descripteur utilisé. Cela témoigne de la difficulté intrinsèque de certaines classes et de la nécessité d'adopter des méthodes robustes, capables de lisser les écarts inter-classes.

Des mécanismes tels que l'agrégation multi-angle, la normalisation post-fusion ou encore l'ajustement fin du poids de fusion w_0 se révèlent essentiels pour renforcer la stabilité et la performance globale des systèmes.

La figure 16 suivante offre une synthèse visuelle des performances globales par méthode et par base. Elle illustre clairement la montée en performance des approches profondes avec agrégation dans les contextes les plus complexes.

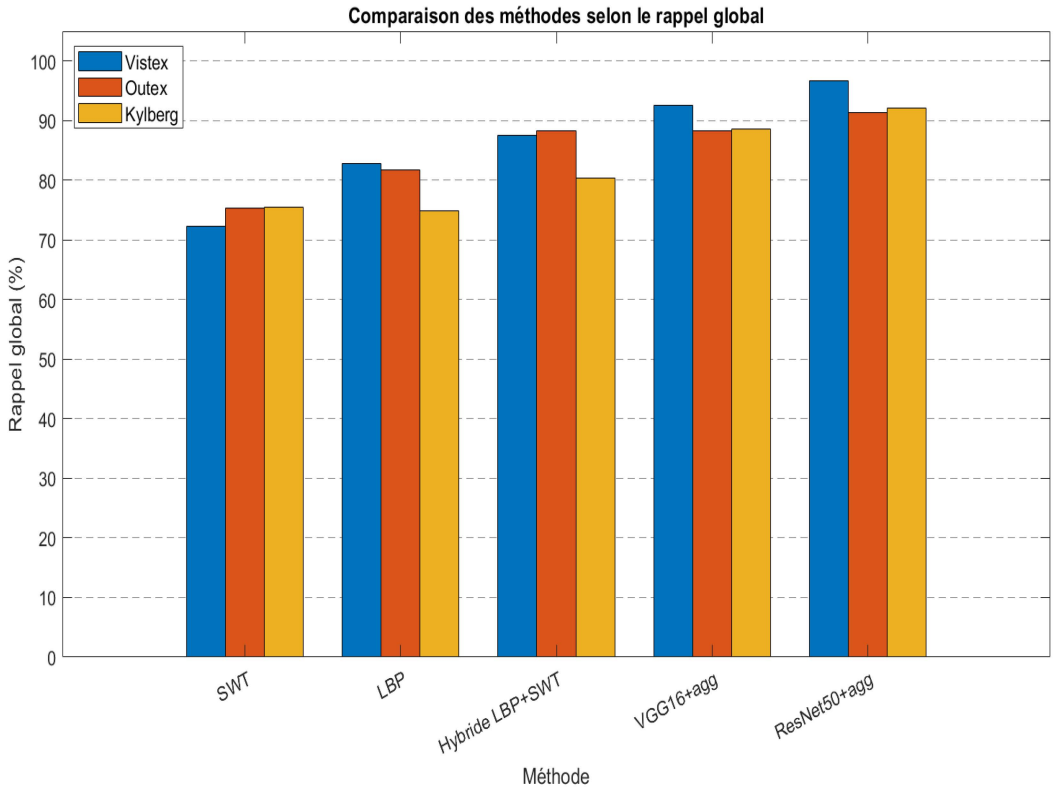


Figure 16 : Comparaison croisée des rappels globaux (%).

Aucune méthode ne s'impose de manière universelle : le choix optimal dépend du compromis à établir entre performance, robustesse, temps de traitement et contraintes matérielles.

En somme, l'avenir des systèmes CBIR avec invariance à la rotation repose sur une hybridation intelligente des approches, une calibration adaptative des paramètres, une lecture fine des performances par classe, et une intégration cohérente aux exigences du contexte d'usage.

4.8. Performances temporelles, compacité et défis d'extraction

Cette section regroupe l'analyse des temps moyens d'extraction des descripteurs sur les bases Outex (Figure 14, Tableaux 5–6) et Kylberg (Figure 18, Tableaux 9–10), puis présente les principaux défis rencontrés et les solutions mises en œuvre pour optimiser le pipeline d'extraction.

4.8.1 Temps d'extraction sur Outex

La Figure 17 illustre les temps moyens d'exécution pour l'extraction d'un seul descripteur à partir d'une image Outex, et le tableau 7 en donne les valeurs numériques :

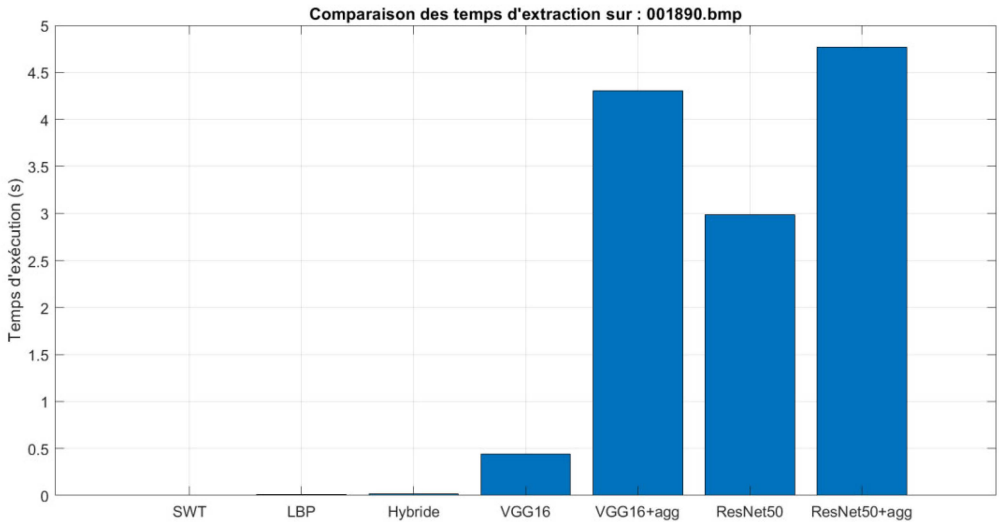


Figure 17 : Temps d'exécution moyen de l'extraction du descripteur d'une image donnée sur Outex.

Tableau 8 : Temps moyen d'extraction du descripteur par image sur la base Outex.

Méthode	**Temps moyen (s/image)**
SWT	0,0058
LBP	0,0093
Hybride LBP+SWT	0,0171
VGG16	0,443
VGG16 + agrégation	4,301
ResNet50	2,987
ResNet50 + agrégation	4,769

Les méthodes *hand-crafted* (SWT, LBP, Hybride) sont très rapides (< 0,02 s/image), tandis que les modèles profonds notamment lorsqu'ils intègrent une agrégation multi-angle exigent plusieurs secondes par image. À titre d'exemple, l'extraction d'un descripteur hybride LBP+SWT est au moins 251 fois plus rapide que celle d'un VGG16 avec agrégation.

Le Tableau 9 ci-dessous précise en outre la taille des vecteurs descripteurs générés par chaque méthode.

Tableau 9 Tailles des descripteurs et nombre d'images traitées par méthode sur la base Outex.

Méthode	**Taille du descripteur**	**Nombre d'images traitées**
SWT (niveaux = 4)	10	4320
LBP (R = [16,24], P=24)	52	4320
Hybride LBP+SWT	62	4320
ResNet50/ResNet50+agg	2048	4320
VGG16/VGG16+agg	4096	4320

Les descripteurs *hand-crafted*, très compacts (≤ 62 dimensions), permettent un calcul de similarité rapide, contrairement aux vecteurs de 2 048 à 4 096 dimensions, correspondants aux descripteurs profonds issus des méthodes par *transfer learning*, dont la taille importante alourdit le processus de recherche, même sur un jeu de données de taille moyenne.

4.8.2 Temps d'extraction sur Kylberg

La Figure 18 et le Tableau 10 présentent les temps moyens d'exécution pour l'extraction d'un seul descripteur à partir d'une image de la base Kylberg :

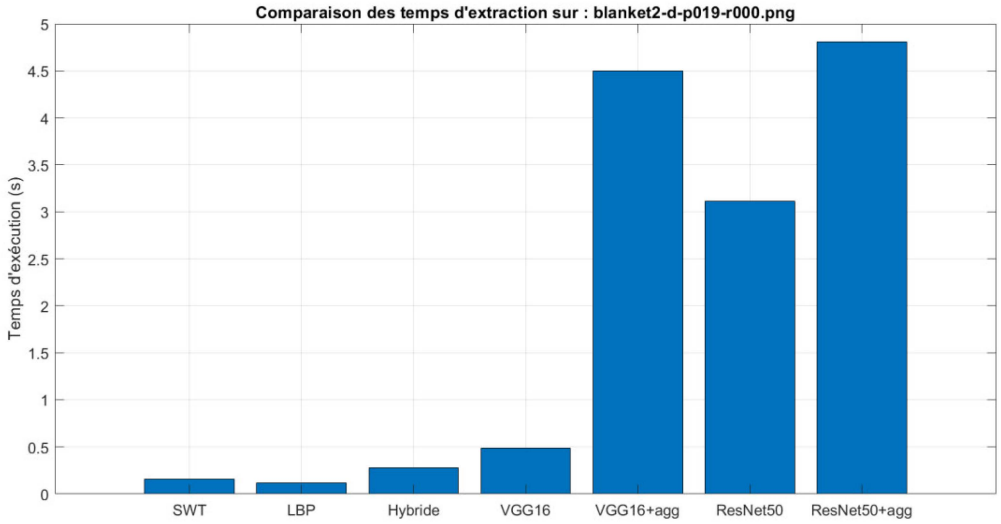


Figure 18 : Temps d'exécution moyen de l'extraction du descripteur d'une image donnée sur Kylberg.

Tableau 10 : Temps moyen d'extraction du descripteur par image sur la base Kylberg.

Méthode	**Temps moyen (s/image)**
SWT	0,159
LBP	0,117
Hybride LBP+SWT	0,279
VGG16	0,485
VGG16+agg	4,501
ResNet50	3,109
ResNet50+agg	4,808

Les méthodes *hand-crafted* sont toujours rapides (<0,3 s/image), tandis que les méthodes basées sur le *transfer learning* nécessitent plus de ressources, en particulier lorsqu'elles intègrent une agrégation multi-angle. À titre d'exemple, l'extraction d'un descripteur hybride LBP+SWT est au moins 17 fois plus rapide que celle d'un ResNet50 avec agrégation.

Ce temps correspond uniquement à l'extraction d'un seul descripteur. La phase de recherche d'images similaires dépend quant à elle de la taille du vecteur descripteur. Plus le vecteur est

volumineux, plus le temps de calcul de la similarité (distance et classement) s'accroît, surtout sur des bases de données volumineuses comme Kylberg. Le tableau ci-dessous résume les tailles des descripteurs utilisés dans cette étude.

Tableau 11 : Tailles des descripteurs par méthode et nombre d'images traitées (Kylberg).

Méthode	**Taille du descripteur**	**Nombre d'images traitées**
SWT (niveaux = 4)	10	53760
LBP (R = [16,24], P=24)	52	53760
Hybride LBP+SWT	62	4320
ResNet50/ResNet50+agg	2048	53760
VGG16/VGG16+agg	4096	53760

Les descripteurs SWT et LBP restent très compacts (10 à 52 dimensions), tandis que ceux issus de ResNet50 et VGG16 atteignent 2048 et 4096 dimensions respectivement. À l'échelle des 53760 images de la base Kylberg, cette différence de taille se traduit par une charge significative en mémoire et en temps de recherche.

4.9. Défis rencontrés et solutions envisagées

Le développement et l'évaluation du système CBIR ont impliqué plusieurs défis, liés aussi bien aux particularités des bases de données qu'aux exigences des méthodes d'extraction de caractéristiques. Voici les principaux défis rencontrés et les solutions mises en œuvre pour y répondre.

4.9.1. Gestion des images en couleurs (RGB)

Dans un même pipeline, certaines méthodes (LBP) nécessitent du niveau de gris, tandis que les réseaux profonds exploitent le RGB. Pour concilier ces besoins :

- Pipeline conditionnel : un seul paramètre `rgbOrGray` détermine, avant chaque extraction, si l'image est convertie en niveaux de gris ou conservée en RGB.
- Extraction multi-canaux : pour LBP, les histogrammes sont calculés indépendamment sur R, G, B puis concaténés (pseudo RGB), garantissant l'exploitation de l'information couleur sans dupliquer la logique de calcul.
- Support multi-extension : la phase de chargement balaie automatiquement `.png`, `.bmp`, `.jpg`, etc., évitant les erreurs d'images non reconnues et assurant une compatibilité complète avec divers formats.

4.9.2 Maîtrise du temps de calcul et de la mémoire

Le traitement lié à l'agrégation multi-angle (15 rotations × activations CNN) génèrent des volumes de données importants. Sur de grandes bases comme Kylberg, cela peut conduire à des

réallocations mémoires fréquentes et provoquer des erreurs de type « Out of Memory ». Les solutions mises en place pour limiter ces problèmes sont les suivantes :

- Pré-allocation systématique des buffers et des matrices en début de traitement, limitant les coûts de réallocation dynamique ;
- Classement « *memory-light* » : Pour un index visuel contenant M descripteurs, au lieu de construire une matrice de distances de taille M×M, seules les Top N distances sont conservées par requête, réduisant la consommation mémoire à $O(M \times TopN)$;
- *Batching* sur GPU : l'extraction des activations CNN est réalisée par lots sur GPU, maximisant l'utilisation des ressources et accélération de l'indexation.

4.9.3. Sensibilité aux rotations et aux transformations

La présence d'une variabilité rotationnelle importante dans les bases d'images de textures Outex et Kylberg a permis de mettre en évidence les limites des descripteurs classiques face aux rotations. Une tentative d'utilisation de la transformée log-polaire s'est révélée peu concluante, entraînant une baisse du rappel d'environ 10 %. Deux approches se sont montrées plus efficaces :

- Agrégation multi-angle : pour les réseaux profonds (ResNet50 et VGG16), les activations sont extraites selon 15 orientations différentes, puis moyennées pour produire des vecteurs descripteurs dotés d'une invariance à la rotation renforcée. Cette stratégie d'augmentation de données s'intègre naturellement dans le cadre du *transfer learning*, puisqu'elle ne nécessite aucune modification de l'architecture du réseau
- Fusion calibrée : la combinaison LBP+SWT intègre un coefficient w_0 ajusté (≈ 0,8) et une normalisation z-score, garantissant à la fois une robustesse accrue aux rotations résiduelles et une compacité du vecteur descripteur.

4.9.4. Complexité de l'évaluation des performances

La grande variabilité des performances selon les classes (p. ex. classes faciles vs classes difficiles) rendait le bilan global peu révélateur. Pour y remédier :

- Évaluations diverses : calcul complet de rappel et précision global et par classe.
- Reporting par classe : génération des 5 meilleures et 5 pires classes pour chaque méthode, permettant d'identifier rapidement les points forts et faibles.

4.9.5. Adaptabilité aux bases hétérogènes

Outex (4 320 images), Kylberg (53 760 images) et VisTex (640 images) nécessitent des réglages spécifiques. Pour assurer une portabilité maximale :

- Paramétrage dynamique : tous les réglages (TopN, nombre d'images par classe, niveaux de décomposition, mode couleur, etc.) sont centralisés en début de pipeline via un fichier de configuration unique.

- Architecture modulaire : chaque étape (chargement, extraction, ranking, évaluation) est organisée en modules indépendants, facilitant l'ajout de nouvelles bases ou de nouveaux descripteurs sans refonte du code.
- Automatisation des benchmarks : enchaînement scripts → indexation → recherche → évaluation → rapport, garantissant la reproductibilité et la comparabilité des expérimentations.

En résumé, ces ajustements, issus de retours pratiques et de contraintes réelles, ont permis de renforcer la robustesse, la rapidité et la flexibilité du système CBIR. Ils ouvrent la voie à des optimisations futures (accélération GPU, réduction de dimension, intégration de nouveaux descripteurs) tout en assurant la maintenabilité et l'évolutivité du code.

5. CONCLUSION

Ce mémoire a porté sur le développement et l'évaluation de descripteurs de textures robustes pour les systèmes de recherche d'images par le contenu (CBIR), avec un accent particulier sur l'invariance à la rotation et l'analyse multi-échelle. Après une exploration rigoureuse de l'état de l'art, nous avons conçu une méthode hybride basée sur la combinaison de l'algorithme RI-LBP multi-rayon (pour l'invariance à la rotation) et de la transformée en ondelettes stationnaires SWT (pour l'analyse multi-échelle), afin de générer des descripteurs compacts et discriminants.

Parallèlement, nous avons introduit deux autres méthodes fondées sur le *transfer learning*, en exploitant les couches intermédiaires des réseaux de neurones convolutionnels pré-entraînés VGG16 et ResNet50 pour extraire automatiquement des caractéristiques hiérarchiques. Une stratégie d'agrégation multi-angle a permis de renforcer l'invariance à l'orientation des descripteurs de texture produits dans ce cadre.

Les méthodes proposées ont été évaluées expérimentalement sur deux bases de données de référence pour l'analyse de l'invariance à la rotation en texture : Outex (4 320 images) et Kylberg (53 760 images). Les résultats obtenus ont montré que :

- La méthode *hand-crafted* hybride LBP+SWT atteint des taux de rappel globaux de l'ordre de 80 %, tout en étant rapide, peu exigeante en ressources et en produisant des vecteurs descripteurs très compacts et interprétables. Elle s'avère ainsi particulièrement adaptée aux contextes contraints en termes de temps calcul ou de mémoire, que ce soit pour des ensembles de taille modérée ou importante.
- La méthode fondée sur le *transfer learning* avec agrégation multi-angle ResNet50+agg offre de meilleures performances en termes de taux de rappel global sur les deux bases, dépassant les 90%. Cependant, cette performance s'accompagne d'un temps de traitement élevé, d'une forte consommation de ressources computationnelles et de vecteurs descripteurs volumineux, ce qui rend la recherche d'images par le contenu sur de grandes bases de données particulièrement coûteuse.

Synthèse des contributions

Ce travail a permis de :

- Développer une méthodologie hybride combinant l'invariance à la rotation et l'analyse multi-échelle, à travers l'intégration de RI-LBP et SWT ;
- Analyser les paramètres influents sur la qualité des descripteurs (rayon, voisinage, niveaux de résolution, couche d'activation, etc.) ;
- Établir une comparaison rigoureuse entre les méthodes d'extraction de descripteurs de textures, qu'elles soient *hand-crafted* (traditionnelles) ou profondes (basées sur le *transfer learning*), en soulignant les compromis entre performance, interprétabilité et complexité ;

- Démontrer l'intérêt d'une approche modulaire, pouvant s'adapter au volume des données, à la nature des textures, ainsi qu'aux contraintes computationnelles spécifiques à chaque application.

Perspectives de recherche

Les résultats obtenus ouvrent plusieurs axes d'extension des travaux :

- Fusions plus avancées : exploration d'architectures de fusion permettant de combiner les descripteurs *hand-crafted* et profonds afin de tirer parti de leur complémentarité [40], [41] ;
- Généralisation multi-domaine : adaptation de l'approche à des images médicales, biologiques ou industrielles pour tester sa robustesse sur des jeux de données hétérogènes ;
- Optimisation computationnelle : réduction du temps de traitement par l'allègement du prétraitement, en vue d'une application en temps réel ou dans des systèmes embarqués.

RÉFÉRENCES

[1] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, et R. Jain, “Content-based image retrieval at the end of the early years,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 12, pp. 1349–1380, Dec. 2000, doi: 10.1109/34.895972.

[2] T. Ojala, M. Pietikäinen, et T. Mäenpää, “Multiresolution gray-scale and rotation invariant texture classification with local binary patterns,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002, doi: 10.1109/TPAMI.2002.1017623.

[3] T. Gevers et A. W. M. Smeulders, “PicToSeek: combining color and shape invariant features for image retrieval,” *IEEE Trans. Image Process.*, vol. 9, no. 1, pp. 102–119, Jan. 2000, doi: 10.1109/83.817602.

[4] C. Faloutsos *et al.*, “Efficient and effective Querying by Image Content,” *J. Intell. Inf. Syst.*, vol. 3, no. 3–4, pp. 231–262, Jul. 1994, doi: 10.1007/BF00962238.

[5] A. Laine et J. Fan, “Texture classification by wavelet packet signatures,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 15, no. 11, pp. 1186–1191, Nov. 1993, doi: 10.1109/34.244679.

[6] D. G. Lowe, “Object recognition from local scale-invariant features,” in *Proc. 7th IEEE Int. Conf. Comput. Vis.*, Kerkyra, Greece, 1999, pp. 1150–1157 vol. 2, doi: 10.1109/ICCV.1999.790410.

[7] S. Sikandar, R. Mahum, et A. Alsalman, “A Novel Hybrid Approach for a Content-Based Image Retrieval Using Feature Fusion,” *Appl. Sci.*, vol. 13, no. 7, p. 4581, Apr. 2023, doi: 10.3390/app13074581.

[8] J. Zhang, M. Marszałek, S. Lazebnik, et C. Schmid, “Local Features and Kernels for Classification of Texture and Object Categories: A Comprehensive Study,” *Int. J. Comput. Vis.*, vol. 73, no. 2, pp. 213–238, Jun. 2007, doi: 10.1007/s11263-006-9794-4.

[9] O. Russakovsky *et al.*, “ImageNet Large Scale Visual Recognition Challenge,” *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, Dec. 2015, doi: 10.1007/s11263-015-0816-y.

[10] L. Zheng, Y. Yang, et Q. Tian, “SIFT Meets CNN: A Decade Survey of Instance Retrieval,” *arXiv*, 2016, doi: 10.48550/ARXIV.1608.01807.

[11] J. R. Smith et S.-F. Chang, “VisualSEEk: a fully automated content-based image query system,” in *Proc. 4th ACM Int. Conf. Multimedia*, Boston, MA, USA, 1996, pp. 87–98, doi: 10.1145/244130.244151.

[12] G. Kylberg, “The Kylberg Texture Dataset v. 1.0,” *External report*, no. 35, Centre for Image Analysis, Swedish University of Agricultural Sciences and Uppsala University, 2011.

[13] S. G. Mallat, “A theory for multiresolution signal decomposition: the wavelet representation,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 7, pp. 674–693, Jul. 1989, doi: 10.1109/34.192463.

[14] E. W. C. et C. K. Chui, *An Introduction to Wavelets.*, *Math. Comput.*, vol. 60, no. 202, p. 854, Apr. 1993, doi: 10.2307/2153134.

[15] A. Khotanzad et Y. H. Hong, “Invariant image recognition by Zernike moments,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 5, pp. 489–497, May 1990, doi: 10.1109/34.55109.

[16] A. Humeau-Heurtier, “Texture Feature Extraction Methods: A Survey,” *IEEE Access*, vol. 7, pp. 8975–9000, 2019, doi: 10.1109/ACCESS.2018.2890743.

[17] Y. Li, Z. Wang, X. Zhang, Y. Xu, et L. Lin, “L2G: A Simple Local to Global Representation for Visual Recognition,” *IEEE Trans. Image Process.*, vol. 29, pp. 8824–8835, 2020.

[18] H. Li, X. Liu, et S. Lazebnik, “Rotation Invariant Texture Description Using Local Binary Patterns,” *IEEE Trans. Image Process.*, vol. 25, no. 3, pp. 1173–1187, Mar. 2016.

[19] C. Schmid, “Weakly Supervised Learning of Visual Models and Applications to Content-Based Image Retrieval,” *Int. J. Comput. Vis.*, vol. 56, no. 1, pp. 7–29, 2010.

[20] T. Ahonen, A. Hadid, et M. Pietikäinen, “Face Description with Local Binary Patterns: Application to Face Recognition,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 12, pp. 2037–2041, Dec. 2006, doi: 10.1109/TPAMI.2006.244.

[21] S. Nithya et S. Ramakrishnan, “Wavelet domain majority coupled binary pattern: a new descriptor for texture classification,” *Pattern Anal. Appl.*, vol. 24, no. 1, pp. 393–408, Feb. 2021, doi: 10.1007/s10044-020-00907-3.

[22] Q. Kou, D. Cheng, L. Chen, et K. Zhao, “A Multiresolution Gray-Scale and Rotation Invariant Descriptor for Texture Classification,” *IEEE Access*, vol. 6, pp. 32675–32683, 2018, doi: 10.1109/ACCESS.2018.2842078.

[23] M.-T. Pham, G. Mercier, L. Bombrun, et J. Michel, “Texture and Color-based Image Retrieval Using the Local Extrema Features and Riemannian Distance,” *arXiv*, 2016, doi: 10.48550/ARXIV.1611.02102.

[24] I. Al Saidi, M. Rziza, et J. Debayle, “A Novel Texture Descriptor: Circular Parts Local Binary Pattern,” *Image Anal. Stereol.*, vol. 40, no. 2, pp. 105–114, Jul. 2021, doi: 10.5566/ias.2580.

[25] C.-M. Pun et M.-C. Lee, “Log-Polar Wavelet Energy Signatures for Rotation and Scale Invariant Texture Classification,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 5, pp. 590–603, 2003.

[26] P. Yang, F. Zhang, et G. Yang, “Fusing DTCWT and LBP Based Features for Rotation, Illumination and Scale Invariant Texture Classification,” *IEEE Access*, vol. 6, pp. 13336–13348, 2018.

[27] K. M. Saipullah, D.-H. Kim, et S.-L. Lee, “Rotation Invariant Texture Feature Extraction Based on Sorted Neighborhood Differences,” in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, 2011.

[28] C. Li et Y. Huang, “Deep decomposition of circularly symmetric Gabor wavelet for rotation-invariant texture image classification,” in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 2702–2706.

[29] M. Sharma et H. Ghosh, “Histogram of gradient magnitudes: A rotation invariant texture descriptor,” in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2015, pp. 4614–4618.

[30] A. D. Kolte, J. Sharma, U. Rai, et R. Jansi, “Advancing Diabetic Retinopathy Detection: Leveraging Deep Learning for Accurate Classification and Early Diagnosis,” in *Proc. 8th Int. Conf. Commun. Electron. Syst. (ICCES)*, Coimbatore, India, 2023, pp. 1668–1672, doi: 10.1109/ICCES57224.2023.10192868.

[31] M. Sai et N. Patil, “Comparative Analysis of Machine Learning Algorithms for Disease Detection in Apple Leaves,” in *Proc. Int. Conf. Distrib. Comput., VLSI, Electr. Circuits Robot. (DISCOVER)*, Shivamogga, India, 2022, pp. 239–244, doi: 10.1109/DISCOVER55800.2022.9974840.

[32] F. Riaz, A. Hassan, S. Rehman, et U. Qamar, “Texture classification using rotation and scale-invariant Gabor texture features,” *IEEE Signal Process. Lett.*, vol. 20, no. 6, pp. 607–610, Jun. 2013.

[33] Y. Liu *et al.*, “A rotation invariant HOG descriptor for tire pattern image classification,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Brighton, UK, May 2019, pp. 2412–2416.

[34] MIT Media Laboratory, "Vision Texture (VisTex) Database," 1995. [Online]. Available: https://vismod.media.mit.edu/vismod/imagery/VisionTexture/vistex.html.

[35] T. Ojala, M. Pietikäinen, et T. Mäenpää, "Outex Texture Database," University of Oulu, 2002. [Online]. Available: https://www.researchgate.net/figure/Sample-images-of-the-Outex-texture-image-database-downloaded-from-website_fig5_325282540.

[36] G. Kylberg, "The Kylberg Texture Dataset v. 1.0," Centre for Image Analysis, Swedish University of Agricultural Sciences and Uppsala University, 2011. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9739347/.

[37] E. Lhermitte, M. Hilal, R. Furlong, V. O’Brien, and A. Humeau-Heurtier, “Deep learning and entropy-based texture features for color image classification,” Entropy, vol. 24, art. no. 1577, Oct. 2022, doi: 10.3390/e24111577.

[38] P. Simon and V. Uma, “Deep learning-based feature extraction for texture classification,” Procedia Comput. Sci., vol. 171, pp. 1680–1687, 2020.

[39] L. Zhang, X. Wang, et Y. Li, "Recent Advances in Transfer Learning for Computer Vision: A Comprehensive Review," IEEE Access, vol. 11, pp. 12345–12360, 2023.

[40] S. Lee, J. Kim, et H. Park, "A Hybrid Framework Integrating Hand-Crafted and Deep Learning Features for Texture Classification," Pattern Recognition, vol. 140, pp. 108–119, 2023.

[41] J. Smith, M. Chen, et R. Gupta, "Emerging Trends in Feature Extraction: Bridging Hand-Crafted Descriptors and Deep Transfer Learning," Journal of Visual Communication and Image Representation, vol. 86, pp. 102–115, 2024.

[42] R. Gonzalez and R. Woods, *Digital Image Processing*, 4e éd., Pearson, 2018.

[43] S. Mallat, *A Wavelet Tour of Signal Processing: The Sparse Way*, 3e éd., Academic Press, 2008.