



Exploitation d'une information multiéclairages pour une approche générique de l'inspection automatique de la qualité visuelle des produits en industrie.

Thomas Lacombe

► To cite this version:

Thomas Lacombe. Exploitation d'une information multiéclairages pour une approche générique de l'inspection automatique de la qualité visuelle des produits en industrie.. Machine Learning [stat.ML]. Communauté Université Grenoble Alpes, 2018. Français. NNT: . tel-01959270

HAL Id: tel-01959270

<https://hal.science/tel-01959270>

Submitted on 18 Dec 2018

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

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

THÈSE

Pour obtenir le grade de

DOCTEUR DE LA COMMUNAUTÉ UNIVERSITÉ GRENOBLE ALPES

Spécialité : **Sciences pour l'Ingénieur**

Arrêté ministériel : 25 mai 2016

Présentée par

Thomas LACOMBE

Thèse dirigée par **Maurice PILLET**
et codirigée par **Hugues FAVRELIERE**

préparée au sein du **Laboratoire SYMME : SYstèmes et Matériaux pour
la MÉcatronique**
dans **l'École Doctorale SISEO : Sciences et Ingénierie des Systèmes,
de l'Environnement et des Organisations**

Exploitation d'une information multi- éclairages pour une approche générique de l'inspection automatique de la qualité visuelle des produits en industrie.

Thèse soutenue à huis clos le **27 Septembre 2018**,
devant le jury composé de :

M. Patrick CHARPENTIER

Professeur - Université de Lorraine - CRAN (Centre de Recherche en Automatique de Nancy), Examineur

M. Alamin MANSOURI

Professeur - Université de Bourgogne - Le2i (Laboratoire Electronique, Informatique et Image), Rapporteur

M. Abdessamad KOBİ

Professeur - Université d'Angers - LARIS (Laboratoire Angevin de Recherche en Ingénierie des Systèmes), Rapporteur

M. Maurice PILLET

Professeur - Université Savoie Mont-Blanc - Laboratoire SYMME (Systèmes et Matériaux pour le Mécatronique), Directeur de thèse

M. Hugues FAVRELIERE

Maître de Conférences - Université Savoie Mont-Blanc - Laboratoire SYMME (Systèmes et Matériaux pour le Mécatronique), Co-Directeur de thèse

M. Yannick EDEL

Expert Industriel - Entreprise EDEL TAMP, Invité



*A mes parents,
pour leur amour et leur dévouement à leurs enfants,
qui m'ont permis d'arriver jusqu'ici*

Remerciements

Une thèse est un travail personnel et collectif à la fois, qui permet de mener un travail réflexif sur un sujet de recherche, mais également sur soi-même. De nombreuses personnes ont joué un rôle dans le parcours qui m'a amené jusqu'ici. Je veux les remercier ici.

Je souhaite tout d'abord remercier l'entreprise Edel Tamp et le Conseil Savoie Mont Blanc (anciennement Assemblée des Pays de Savoie) qui, par leur partenariat de co-financement, ont rendu possible la réalisation de ces travaux. Le projet SPOT qui en est né est un bel exemple de partenariat public/privé fertile en terme d'idées et menant à des développements applicatifs d'intérêt.

Je souhaite également remercier M. Alamin Mansouri et M. Abdessamad Kobi pour avoir accepté de relire ce travail de thèse et à M. Patrick Charpentier pour avoir accepté le rôle d'examineur.

Un grand merci à mon directeur de thèse, Maurice Pillet, pour m'avoir choisi et fait confiance dans la conduction de ces travaux de thèse. Ses conseils toujours avisés, son expérience et son dynamisme en toutes circonstances m'ont beaucoup appris et fait évoluer tout au long de ces trois années, autant sur le plan scientifique que sur le plan personnel.

Un autre grand merci à mon co-directeur de thèse, Hugues Favrelière, pour son soutien tout au long de ces trois ans. Toujours présent et réactif lorsque j'en ais eu besoin, ce fut un plaisir de partager ces trois années de travail au laboratoire.

Merci à M. Yannick Edel, président-directeur d'Edel Tamp, pour son soutien, son engagement et sa réactivité dans le cadre du projet. Merci également pour son entrain et sa pédagogie lors des visites des sites de production et pour son accueil avec toutes les équipes d'Edel Tamp lors de mes quelques passages dans les locaux de l'entreprise. Un remerciement spécial à Patrick Oriol avec qui j'ai particulièrement apprécié travailler sur la conception de la machine SPOT.

Merci à Laurent Tabourot, directeur du laboratoire SYMME pendant la période de ma thèse, et directeur de mon stage de dernière année d'école d'ingénieur, pour le partage de son expérience. Son intuition quant à mon attrait pour la recherche et à mes capacités à me lancer dans un travail de thèse s'est révélée juste, et je le remercie de m'avoir orienté dans cette voie. Merci également à Antoine Machut et Thierry Gerez pour leur retour, ressentis et conseils quant à leur propre expérience de thèse, qui m'ont aidé à prendre la décision de me lancer !

Pour remonter encore un peu plus loin, je souhaite aussi remercier M. Tristan Rondepierre, qui fut mon professeur de physique-chimie en terminale, et qui tient un rôle important dans la passion que j'ai développée pour les sciences et pour l'enseignement. Le partage de sa propre passion pour les sciences, et notamment l'astronomie au travers des séances d'observations organisées à l'observatoire de Lyon, et son dévouement pour ses élèves m'ont inspiré et m'inspirent toujours pour poursuivre dans l'enseignement et la recherche.

Un très grand merci à tous les personnels du laboratoire SYMME pour leur accueil et les moments partagés lors de ces 3 ans. Je n'ai pas peur de dire que je me sens au laboratoire comme

dans une seconde maison, et que la bienveillance et la gentillesse de chacun y est pour beaucoup.

Parmi toutes ces personnes, je souhaite particulièrement remercier Françoise, Jérôme, Nasika et Jean-Christophe, dont le travail et l'engagement rendent la vie au laboratoire agréable pour chacun et chacune. Leurs efforts de tous les jours ne sont pas forcément toujours mis en avant, mais je souhaite le faire ici car ils le méritent.

Parmi les personnels du laboratoire, je souhaite également remercier tous les enseignants-chercheurs (je ne citerais personne car j'en oublierai sûrement) avec qui j'ai été en contact (notamment autour de la pause café sacrée de 10h) et avec lesquels j'ai toujours eu beaucoup de plaisir à échanger.

Un autre très grand merci à tous les doctorants que j'ai pu côtoyer au laboratoire, dont certains et certaines sont devenu de très bons amis et amies ! J'ai eu de la chance d'intégrer le laboratoire en même temps qu'un groupe dynamique d'autres doctorants, avec lesquels nous avons eu l'occasion de construire des souvenirs impérissables.

Merci plus particulièrement à Amandine pour sa gentillesse et sa générosité, son aide avec ~~ATE~~ lors des derniers instant de rédaction, mais aussi les parties de ping-pong intenses et les initiations au karaté.

Merci à Thomas dit "M", pour cette belle amitié, également pour les intenses parties de ping-pong, mais aussi ces moments partagés avec le groupe international, cette courte période de collocation et cette prose si élégante qui est la sienne.

Merci à Thomas dit "coach" et Orlando, pour la belle collocation entre doctorants et futurs doctorants symmiens.

Merci à Bruno, également pour cette belle amitié, et cette collaboration fructueuse entre visuel et haptique ! J'ai vraiment été touché de pouvoir assister à ton mariage, merci encore de l'invitation. Je peux continuer comme cela avec chaque autre doctorant du SYMME tellement j'ai de souvenir à raconter, mais, j'espère qu'ils m'excuseront, je vais seulement lister tous les autres. Merci donc à Jérémy (St Patrick's power), Mathias U. et Mathias C., Pierre (merci aussi pour l'invitation à ton mariage), Amélie, Thibault, Aurélien, Jean, Soumaya et tous les autres !

Merci aussi aux plus anciens qui étaient là à mon arrivé mais qui sont partis depuis, Sylvain, Laurent, Florian, Léonida, Gilles, Gwyddyon ou encore Yohan. Un remerciement spécial à Simon qui m'a repéré pour ce sujet de thèse et qui m'a ensuite tutoré dans mon début de parcours.

Merci également aux doctorants des autres laboratoires du campus d'Annecy avec qui j'ai pu nouer de belles amitiés et qui m'ont aidé avec motivation à renforcer les liens inter-laboratoires ! Merci donc à Annaïg, Emilie, Boris, Alexandre, Mériem, Quentin, Mathilde, Philippe, Alice et tous les autres !

Je veux aussi remercier les collègues de l'IUT d'Annecy et plus particulièrement du département Mesures Physiques où j'ai enseigné pendant deux ans dans le cadre de mon service DCE. Merci particulièrement à Myriam et Guillaume qui m'ont plus spécialement suivi et accompagné dans mon expérience d'enseignement.

Merci à l'équipe des représentants au conseil de l'Ecole Doctorale SISEO et plus particulièrement à Emilie et Boris, pour les aller-retours au Bourget-du-Lac dans le cadre des conseils dans une ambiance toujours agréable (tant qu'on ne débat pas de choses qui fâchent !), et pour les événements "Lâche ta thèse !" dont ce week-end mémorable au Roc des Bœufs !

Merci également à tous les membres du conseil doctoral et plus particulièrement à Emmanuel Trouvé, sous la direction duquel j'ai vraiment apprécié de participer à la vie de l'ED, et aux dif-

férentes secrétaires de l'ED sans qui rien ne tournerait !

Merci à l'équipe du comité opérationnel du projet QVT campus pour la belle aventure que nous avons vécu tous ensemble. J'ai vraiment apprécié de vous rencontrer et de pouvoir casser les frontières entre composantes et services du campus lors de ces réunions et des déjeuners que nous continuons à organiser.

Merci aux étudiants stagiaires et en projet de DUT et d'école d'ingénieur qui m'ont aidé à avancer dans ce projet, notamment Vincent, Quentin et Thomas.

Parmi toutes les activités que nous avons instauré pour décompresser lors de nos temps libres au laboratoire avec les doctorants et les enseignant-chercheurs du laboratoire, je souhaite remercier la team grimpe, Ludo, Thibault, Aurélien, Thomas, Émile, Mathias, Alex, Pierre et Boris. J'ai redécouvert cette fabuleuse activité qu'est l'escalade, en mur, en bloc ou en falaise, avec vous et je vous en suis très reconnaissant.

Merci à tous mes autres amis qui m'ont soutenu de près à Annecy ou de plus loin (à Lyon Grenoble ou ailleurs). Merci donc à Félix, Kelly, Sierra, Thomas "M", Thomas "coach", Orlando, Ivan, toutes les géniales personnes que j'ai rencontré au sein du groupe international à Annecy, mais aussi mes amis de plus longue date, Yann, François, Rémi, Léonie, Axel, Sylvain, Baptiste, Bastien, Élise et tous et toutes les autres.

Merci à la dream team des cousins Marie et Simon et ma sœur Audrey, pour cette relation privilégiée que nous avons depuis si longtemps et pour encore longtemps je l'espère. Particulièrement à ma sœur qui me supporte (dans les différents sens du terme) depuis toujours !

Merci à mes parents, pour leur soutien sans faille dans ma vie et mes choix et leur franchise quand je nécessite leur conseil. Pour leur dévouement à ma sœur et à moi, dont je leur suis si reconnaissant car il m'a permis d'arriver jusqu'ici. Pour leur amour.

Finally, I want to turn to English for my last thanks. To the amazing girl I have met on the middle of this PhD, in a magical situation in Florence. Thank you my Mary-Anne for your incredible support and patience with me during this period. You're the most beautiful discovery I have done during these 3 years. I love you but tons !

Je m'excuse pour toutes les personnes qui ont joué un rôle et qui ont compté pour moi dans la réussite de ce projet de thèse, et que j'aurais oublié dans ces remerciements.

Avant Propos

Ces travaux de thèse ont été réalisés au sein du laboratoire **SY**stèmes, et **Mat**ériaux pour la **ME**catronique (SYMME), localisé à Annecy et faisant parti de l'Université Savoie Mont Blanc. Ils s'inscrivent dans le contexte du projet **Surface PerceptiOn Tolerancing** (SPOT), né du partenariat entre le laboratoire et l'entreprise industrielle EDEL TAMP. Le financement de cette thèse est partagé à parts égales entre l'entreprise partenaire et le Conseil Savoie Mont Blanc (anciennement Assemblée des Pays de Savoie).

Fondée en 1991, l'entreprise EDEL TAMP est spécialisée dans le développement et la production de pièces plastiques injectées. Elle regroupe des expertises multi-métiers lui permettant d'offrir un service allant de la conception jusqu'à l'industrialisation. Elle s'occupe du développement de produits pour des domaines d'applications variés tels que l'agriculture, le médical, le bâtiment, la manutention ou encore le jardinage.

Créé en 2006, le laboratoire SYMME regroupe les compétences de 47 enseignants chercheurs et d'une vingtaine de doctorants repartis sur deux axes de recherches : l'axe Qualité Industrielle (QI) et l'axe Matériaux, Systèmes et Instrumentation Intelligents (MSII). Les différents projets portés dans ces deux axes trouvent principalement des applications dans les domaines de l'Energie, de la Santé et de la Production et Organisation Industrielle. Le projet SPOT s'inscrit dans l'axe QI, et plus particulièrement dans la thématique de la qualité sensorielle.

Cette thématique de la qualité sensorielle a été développée au laboratoire au travers de différents projets et travaux de thèses résumés en Figure 1. Les travaux développés dans cette thèse font suite à différents projets ayant tout d'abord portés sur la modélisation et de la réduction de la variabilité du contrôle visuel, avec notamment les thèses de Anne-Sophie Guerra, Nathalie Baudet et Gaëtan Le Goïc. Les travaux de ce dernier ont également porté sur une évolution vers l'automatisation du contrôle visuel, thématique poursuivie avec les thèses de Simon Desage et Gilles Pitard. D'autres thématiques parallèles se sont également développées, notamment autour de plusieurs méthodes de tolérancement, dont le tolérancement modal développé notamment avec les travaux de thèse de Hugues Favrelière. Plus récemment, cette expertise du contrôle sensoriel s'est également ouverte au domaine de la qualité haptique avec les travaux menés par Bruno Albert en cotutelle avec l'université de Strasbourg.



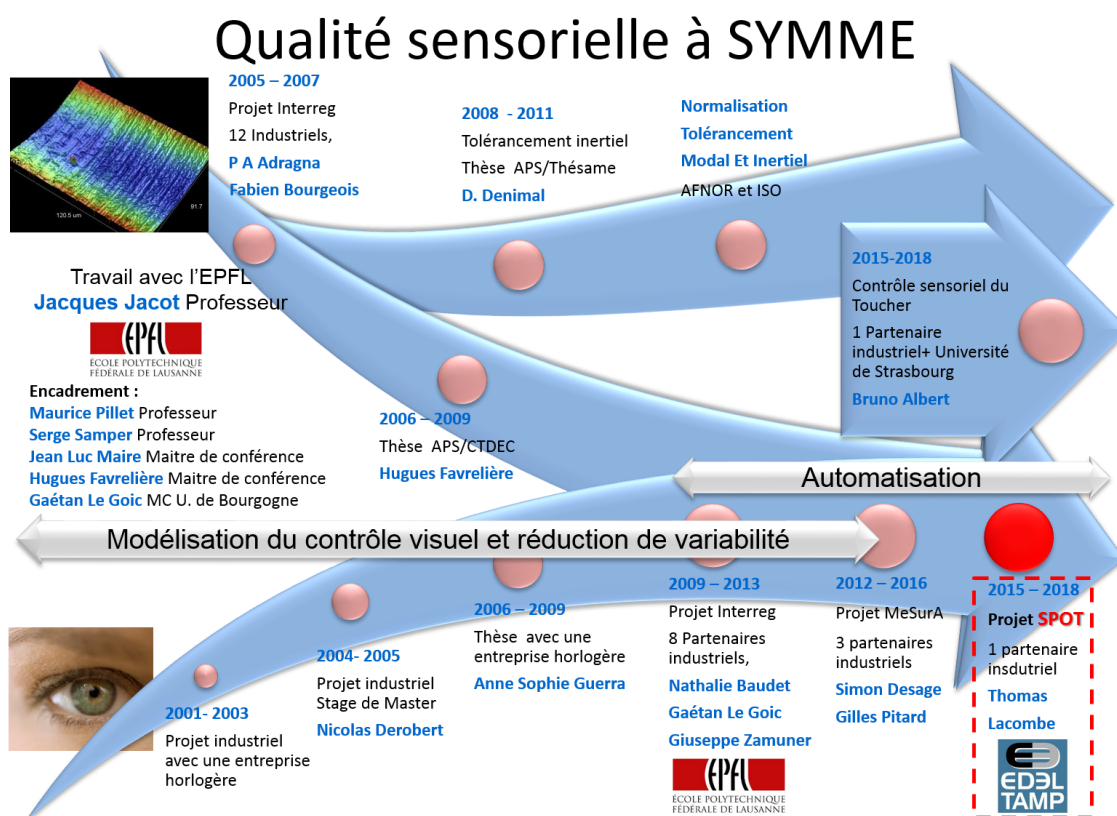


Figure 1 – Historique des projet et thèses menées au laboratoire SYMME autour de la problématique de la qualité sensorielle des produits

Table des matières

Remerciements	i
Avant Propos	v
Table des matières	ix
Introduction générale	1
1 Contrôle visuel d'aspect : d'un contrôle manuel à un contrôle automatisé	7
1.1 Introduction	8
1.2 Le contrôle visuel humain en industrie	10
1.2.1 Perception humaine et variabilité du contrôle visuel	10
1.2.1.1 Perception et sources d'influence	10
1.2.1.2 Variabilité des résultats du contrôle visuel industriel	13
1.2.2 Formaliser le contrôle visuel humain	15
1.2.2.1 Décrire les anomalies	15
1.2.2.2 Modéliser les procédures de contrôle	16
1.2.2.3 Méthodologie d'exploration dans le cadre du contrôle	17
1.2.3 Vers une automatisation du contrôle visuel	19
1.2.4 Synthèse quant au contrôle visuel humain	23
1.3 État des lieux du contrôle automatique par vision industrielle	24
1.3.1 Automatisation et systèmes de vision industrielle	24
1.3.1.1 Classifications des applications des systèmes de vision industrielle	24
1.3.1.2 Points clefs pour la conception d'un système de vision industrielle	29
1.3.2 Positionnement applicatif du contrôle automatisé de la qualité d'aspect	31
1.3.2.1 Secteurs d'activités	31
1.3.2.2 Moyens pour traiter l'information	34
1.3.3 Cahier des charges d'un système de contrôle reproduisant les caractéristiques spécifiques d'un contrôle humain	36
1.4 Conclusion	38
2 Réduction de l'information des images : Description	41
2.1 Introduction	42
2.2 Décrire une image : un état des lieux	43
2.2.1 Approches géométriques ou de forme	43
2.2.1.1 Représentation des images	43
2.2.1.2 Descripteurs géométriques	44
2.2.2 Approches statistiques ou de distributions	45
2.2.2.1 Approches de représentation	45
2.2.2.2 Descripteurs statistiques	50
2.2.3 Approches filtrage ou transformation	52
2.2.3.1 Domaine spatial	52
2.2.3.2 Domaine fréquentiel	52

2.2.3.3	Domaine spatio-fréquentiel	52
2.3	Décomposition modale discrète pour la description d'images	54
2.3.1	Principe de fonctionnement	55
2.3.2	Décomposition d'une géométrie	56
2.3.3	Application à l'analyse de l'aspect visuel des surfaces	58
2.3.4	Analyse multi-échelle par filtrage modal	58
2.3.5	Approximation de la surface de luminance	60
2.3.6	Description d'images via la DMD	62
2.4	Travaux applicatifs utilisant la description d'images	68
2.4.1	La DMD comme un indicateur de performance en Deep Learning	68
2.4.2	Description d'images pour prédire des caractéristiques haptiques	71
2.5	Conclusion	76
3	Réduction de l'information des images : Tri des descripteurs	77
3.1	Introduction	79
3.2	Sélection de descripteurs	80
3.2.1	Intérêt dans le contexte de nos travaux	80
3.2.2	Principes généraux	80
3.2.3	Procédure de génération	82
3.2.3.1	Génération complète ou exhaustive	82
3.2.3.2	Génération séquentielle	84
3.2.3.3	Génération aléatoire	85
3.2.3.4	Aide au choix d'un algorithme	87
3.2.4	Fonction d'évaluation	89
3.2.4.1	Mesures de distances	90
3.2.4.2	Mesures d'information	94
3.2.4.3	Mesures de dépendances	95
3.2.4.4	Mesures de consistance	96
3.2.4.5	Mesures du taux d'erreur de classification	96
3.2.5	Critère d'arrêt	97
3.2.6	Validation	98
3.3	Transformation de l'espace des descripteurs	99
3.3.1	Analyse en Composantes Principales (ACP)	99
3.3.2	Analyse en Composantes Indépendantes (ACI)	101
3.3.3	Analyse Linéaire Discriminante (LDA)	102
3.3.4	Algorithme t-SNE	103
3.3.5	Synthèse et mise en perspective dans la cadre du contrôle visuel automatisé en industrie	103
3.4	Critère de séparation basé sur le test du T^2	104
3.4.1	Test du T^2	104
3.4.1.1	Principes du test statistique	105
3.4.1.2	Statistique de test du T^2	105
3.4.2	Application aux cartes de contrôle	106
3.4.3	Critère d'évaluation de distance par le T^2	107
3.5	Conclusion	110
4	Classification de données	113
4.1	Introduction	115
4.2	Méthodes pour la classification de données	117
4.2.1	k-plus-proches voisins	117
4.2.1.1	Principe de fonctionnement	117

4.2.1.2	Distance entre deux éléments	118
4.2.1.3	Choix de k	118
4.2.1.4	Forces et limites	119
4.2.2	Réseaux bayésiens naïfs	120
4.2.2.1	Théorème de Bayes	120
4.2.2.2	Réseaux bayésiens	121
4.2.2.3	Classification par réseaux bayésiens naïfs	122
4.2.2.4	Forces et limites	123
4.2.3	Machines à Vecteurs de Support - Support Vector Machine (SVM)	124
4.2.3.1	Cas d'une séparation linéaire	124
4.2.3.2	Cas d'une séparation non-linéaire	125
4.2.3.3	Cas d'une classification multi-classes	127
4.2.3.4	Forces et limites	128
4.2.4	Arbres décisionnels et Forêts d'arbres décisionnels	128
4.2.4.1	Classification par arbre de décision	128
4.2.4.2	Forces et limites	132
4.2.4.3	Quelques évolutions des algorithmes d'arbres décisionnels	133
4.2.5	Réseaux de neurones	135
4.2.5.1	Neurone artificiel ou formel	135
4.2.5.2	Perceptron simple	137
4.2.5.3	Perceptron multi-couche	141
4.2.5.4	Forces et limites	144
4.2.6	Algorithme des k-moyennes	145
4.2.6.1	Principe de fonctionnement	145
4.2.6.2	Forces et limites	147
4.2.7	Réseaux de Deep Learning	147
4.2.7.1	Fondements historiques	147
4.2.7.2	Performances d'un réseau de Deep Learning	149
4.2.7.3	Utilisation de réseaux pré-entraînés	150
4.2.7.4	Forces et limites	151
4.3	Discussion quant aux caractéristiques des différentes méthodes et leur application industrielle	152
4.3.1	Comparaison des techniques	153
4.3.2	Mise en perspective dans le contexte du contrôle d'aspect industriel automatisé	157
4.3.3	Méthodes d'ensemble	160
4.3.4	Intérêt des algorithmes non-supervisés	160
4.4	Conclusion générale sur les méthodes de classification	162
5	Application de l'approche multi-éclairages	163
5.1	Introduction	164
5.2	Fonctionnement de l'acquisition des images et du traitement des données	165
5.2.1	Acquisition multi-éclairages	165
5.2.2	Algorithme de traitement des données	165
5.2.2.1	Fonctionnements hors-ligne et en-ligne	165
5.2.2.2	Pré-traitement des images	165
5.2.2.3	Description des images	165
5.2.2.4	Algorithme de réduction de l'ensemble des descripteurs	165
5.2.2.5	Algorithmes de classification	165
5.3	Résultats expérimentaux	165
5.3.1	Conditions expérimentales	165

5.3.2	Résultats de classification	165
5.4	Dispositif d'acquisition d'images SPOT	165
5.4.1	Principes de fonctionnement	165
5.4.2	Synthèse sur les degrés de libertés	165
5.5	Conclusion	166
Conclusion générale et perspectives		168
Bibliographie		I
Table des figures		XVI
Liste des tableaux		XX
A	Annexes	XXII
A.1	Annexe 1 : Exemples de <i>cheat sheets</i> pour la sélection des algorithmes de <i>machine learning</i>	XXIII
A.2	Annexe 2 : Représentation détaillée de la sphère Mesura	XXV

Introduction générale

*Toute science crée une nouvelle
ignorance.*

Henri Michaux
Écrivain et poète
1899 - 1984

Préambule

Maintenir un haut niveau de satisfaction chez ses clients est un défi essentiel pour toute entreprise. Cette satisfaction est directement liée à la notion de qualité des produits proposés, et plus particulièrement à la qualité perçue par les consommateurs au travers de leur perception et de leur ressenti. Cette perception et ce ressenti passant par la capture de signaux aux travers de nos différents sens, on parle alors de qualité sensorielle des produits. La maîtrise de cette dernière constitue donc un enjeu toujours plus important pour les entreprises à l'heure actuelle, dans de très nombreux secteurs d'application.

En effet, si la qualité sensorielle des produits semble originellement capitale dans des domaines d'activité comme le luxe ou encore les cosmétiques, dont la fonctionnalité des produits est fondée sur leur apparence visuelle, la sensation tactile qu'ils procurent ou encore leur odeur par exemple, ce type de critères prend également une place de plus en plus importante dans d'autres domaines d'activités variés. Ainsi, les critères sensoriels s'invitent maintenant dans des cahiers des charges relevant de beaucoup d'autres secteurs, car ils deviennent un critère de qualité globale et confiance dans le produit.

Un exemple parlant peut être pris avec l'industrie agroalimentaire, dans laquelle l'apparence visuelle, l'odeur, le goût et même la sensation tactile des produits dans certains cas jouent un rôle clé dans le choix du consommateur.

Afin de garantir la maîtrise de ces différents aspects, les industriels se dotent de méthodes et processus axés sur le contrôle sensoriel des produits. Dans nos travaux de thèse, nous nous intéressons plus particulièrement au contrôle visuel des produits.

Problématique de recherche

Le contrôle visuel humain en industrie est confronté à des problèmes de variabilité des résultats, dû en partie à l'aspect subjectif de ce type d'inspection. Dans ce contexte, la modélisation du contrôle humain et le développement de procédures d'inspection standardisées ont fait l'objet de recherches, afin de fournir des solutions aux industriels pour réduire la variabilité constatée. En parallèle, des dispositifs de contrôle automatisés ont également vu le jour, afin d'utiliser la capacité des machines à effectuer des routines avec une très bonne répétitivité. Néanmoins, si ces dispositifs sont très performants pour effectuer une tâche spécifique, ils restent encore peu flexibles et souvent développés pour des applications très spécifiques. Il n'existe en effet à l'heure actuelle pas de système capable de détecter tous les défauts sur tous les types de surfaces.

C'est dans ce cadre que s'inscrivent ces travaux de thèse, qui cherchent à proposer une approche plus générique et flexible du contrôle d'aspect automatisé. Pour cela, nous proposons d'adopter une démarche inspirée du contrôle humain en exploitant une information multi-éclairages dans un système de contrôle d'aspect automatisé. Cette idée est née à la confluence de deux constats.

Le premier d'entre-eux est que les systèmes d'inspection automatique industriels actuels exploitent rarement plus d'une image issue soit d'une illumination directionnelle venant d'un angle d'éclairage spécifique, soit d'un éclairage uniforme ambiant.

Le second constat porte sur le fait que le contrôle visuel humain est en général réalisé grâce à une approche utilisant plusieurs angles d'éclairage. En effet, comme nous le faisons souvent pour vérifier l'impeccabilité d'aspect de l'écran de notre téléphone, les contrôleurs en industrie font *miroiter*¹ les pièces pour maximiser les chances de détecter d'éventuelles anomalies à leur surface (rayures, coups, variations de couleur, traces, etc.).

A partir de ces deux constats, nous proposons une approche d'inspection visuelle automatique exploitant une information issue de plusieurs angles d'éclairages, s'inspirant ainsi du processus humain, dans le but de développer des systèmes plus génériques et flexibles en terme de type de pièces et d'anomalies à contrôler.

La thèse défendue par ces travaux postule qu'il est possible d'exploiter une information multi-éclairages pour mieux détecter les anomalies d'aspect variées dans le contexte de l'inspection automatique de la qualité visuelle des produits en industrie. L'application industrielle de ces travaux implique un temps de traitement limité, pour correspondre aux contraintes des lignes de production, notamment dans le domaine de l'injection plastique.

Organisation du manuscrit

Pour argumenter cette thèse, nous proposons d'organiser ce manuscrit en 5 chapitres permettant de décomposer la problématique présentée précédemment. Les chapitres 1, 2, 3 et 4 s'appliquent à apporter des réponses à différentes problématiques présentées dans la Figure 2, suivant l'organisation suivante :

- Chapitre 1 : Acquisition de données inspirée du contrôle humain,
- Chapitres 2 et 3 : Réduction de l'information,

1. Miroiter : faire varier la configuration entre la surface observée, la position de l'œil et l'angle de la lumière incidente.

— Chapitre 4 : Classification de l'information.

Pour finir, le chapitre 5 porte sur la présentation d'une application mettant en avant l'intérêt de l'approche multi-éclairages. Ce dernier chapitre est confidentiel car portant sur des développements qui peuvent mener à des développements industriels et à un dépôt de brevet avec l'entreprise partenaire.

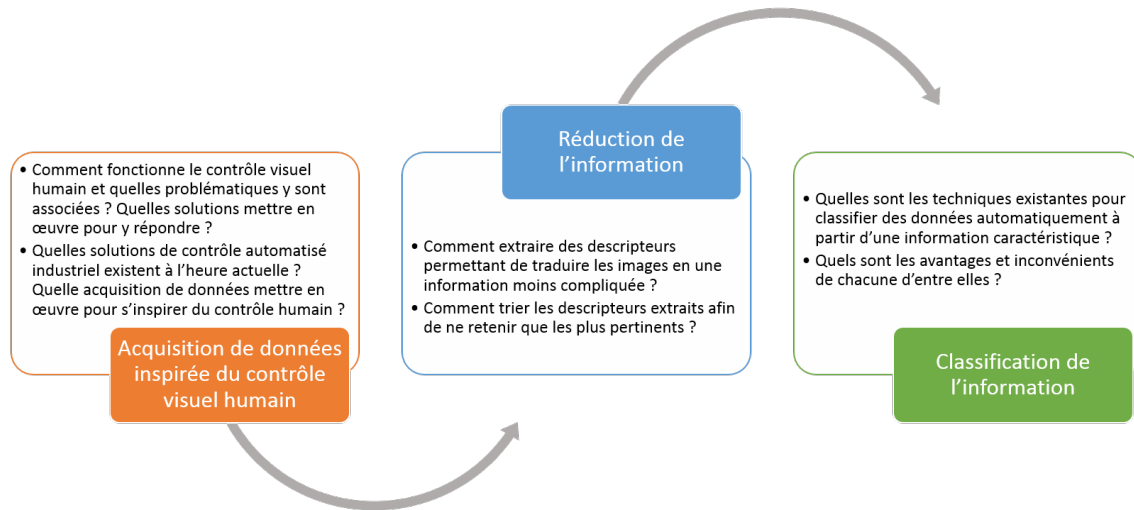


Figure 2 – Problématiques liées au développement d'un système d'inspection d'aspect automatisé inspiré du contrôle visuel humain

Le premier chapitre s'applique tout d'abord à donner un aperçu au lecteur du contexte du contrôle visuel en industrie, afin de mieux comprendre les problématiques ayant motivé ce travail de thèse.

Il s'intéresse dans un premier temps à décrire plus en détails les sources d'influences qui peuvent toucher la perception humaine, et notamment la perception visuelle, afin de mieux comprendre l'apparition d'une variabilité dans le contrôle visuel humain.

Afin de réduire cette variabilité, différents travaux de recherche ont porté sur la modélisation et la formalisation du contrôle visuel humain. Une synthèse bibliographique de ces propositions est ensuite présentée, afin de mieux comprendre ce qu'est une anomalie d'aspect et comment elles peuvent être détectées et évaluées.

A partir de cette modélisation et formalisation, des perspectives d'évolution sont développées pour aller vers une automatisation du contrôle inspirée par le contrôle humain, notamment dans des contextes industriels contraints.

Enfin, ce chapitre se termine par un état des lieux des solutions de vision industrielle déjà existantes. Différentes classifications sont alors présentées en terme de type d'applications, de secteurs d'activité et de techniques utilisées pour ce type de systèmes. Ces différentes considérations mènent au final à la présentation des points clés d'un cahier des charges pour un système de contrôle reproduisant les caractéristiques spécifiques d'un contrôleur humain dans le contexte industriel.

L'approche multi-éclairages fournissant une quantité importante d'information (plusieurs dizaines d'images), il est nécessaire de mettre au point des méthodes pour réduire cette information afin de rendre son exploitation possible sous les contraintes de temps de traitement imposées par le contexte industriel. Les chapitre 2 et 3 sont dédiés à répondre à cette problématique de réduction

de données.

Le chapitre 2 se focalise sur les méthodes permettant de décrire les images, c'est-à-dire d'en extraire des paramètres caractérisant l'information qu'elles contiennent. Les techniques de description d'images permettent de passer d'une image contenant des millions de pixel à un vecteur de valeurs appelées *descripteurs*.

La première partie de ce chapitre est dédiée à une étude bibliographique des méthodes de description d'images les plus communément utilisées pour des applications de classification automatique².

La seconde partie du chapitre s'applique ensuite à introduire l'utilisation d'une nouvelle méthode de description d'images : la Décomposition Modale Discrète (DMD). Cette méthode développée lors des travaux ayant précédé ces travaux de thèse au laboratoire SYMME permet de décomposer une surface dans une base de modes élémentaires. Son application à la description d'images dans un objectif de classification automatique constitue un apport original dans ce travail de thèse. Enfin, nous présentons pour finir d'autres applications de la méthode DMD et de la description d'images dans les contextes de la mesure de performance de génération d'images géométriques et de la prédiction de sensations haptiques. Ces deux applications sont issues de travaux collaboratifs avec d'autres doctorants du laboratoire, et constituent deux apports secondaires de ce travail de thèse.

Même si cette première étape de description des images permet de réduire l'information contenue dans les images, les descripteurs extraits ne sont pas forcément tous pertinents pour garantir une classification efficace. En effet, certains peuvent même dégrader l'information et engendrer une diminution des performances de classification. De plus, l'utilisation d'un trop grand nombre de descripteurs pour l'étape de classification peut mener à des temps de traitement longs et incompatibles avec le contexte industriel qui nous intéresse. Une autre étape de réduction d'information est donc nécessaire pour réduire le nombre de descripteurs issus de l'extraction précédente.

Le chapitre 3 s'applique donc dans un premier temps à exposer les différentes familles de méthodes permettant de réaliser ce type de réduction, en sélectionnant un sous-espace de descripteurs les plus pertinents pour la classification, ou en transformant l'espace des descripteurs en un espace de plus faible dimension. Dans un second temps, nous présentons un nouveau critère de sélection de descripteurs développé au cours de ce travail de thèse et basé sur la statistique du T^2 . La proposition de ce critère constitue un apport original de ce travail de thèse.

Une fois la sélection d'un sous-espace de descripteurs pertinents (pour une classification performante en terme de résultats et de temps de calcul) sélectionnés, ces derniers peuvent être utilisés pour nourrir un algorithme de classification. Cet algorithme, une fois entraîné sur des données d'apprentissage, est capable d'affecter une classe à une nouvelle donnée, dans notre cas une pièce de qualité inconnue produite sur la ligne de production.

Le chapitre 4 s'emploie dans un premier temps à donner un aperçu des principales techniques utilisées pour réaliser ce type de tâche, et d'en expliquer les différents principes de fonctionnement. Les forces et limites de chacun de ces algorithmes sont également présentées afin de pouvoir comparer ces différentes méthodes dans la seconde partie du chapitre. Cette comparaison est réalisée au regard de l'application de la classification automatique dans le contexte industriel qui nous intéresse. Nous souhaitons apporter par cette comparaison et certaines perspectives y étant liées, des pistes pour les développements futurs dans ce domaine.

2. La classification automatique consiste en l'attribution d'une classe à un objet à classer. Dans notre cas, il s'agit par exemple d'attribuer une classe Conforme ou Non Conforme à une pièce contrôlée, dont la qualité est par avance inconnue.

Enfin, l'objet du chapitre 5 porte sur la mise en œuvre des différents concepts présentés précédemment, en proposant une application de contrôle visuel d'aspect multi-éclairages, développée au cours de ces trois années de travail et permettant de justifier l'apport de l'approche multi-éclairages. Les résultats expérimentaux obtenus avec ce système sont présentés afin de mettre en avant l'intérêt de l'approche multi-éclairages pour détecter des anomalies d'aspect, par rapport à une approche n'utilisant qu'une seule image.

Nous présentons également les performances du système en terme de temps de traitement pour justifier son utilisation sous des contraintes industrielles.

Nous terminons enfin en présentant des perspectives de travail quant aux développements présentés dans le reste du chapitre, notamment avec la conception d'une nouvelle machine d'acquisition plus flexible en terme de géométries de pièces observables, et un travail préliminaire mené sur la sélection des points d'éclairages pertinents pour optimiser le contrôle en fonction du type de produit observé.

1

Contrôle visuel d'aspect : d'un contrôle manuel à un contrôle automatisé

Sommaire

1.1	Introduction	8
1.2	Le contrôle visuel humain en industrie	10
1.2.1	Perception humaine et variabilité du contrôle visuel	10
1.2.1.1	Perception et sources d'influence	10
1.2.1.2	Variabilité des résultats du contrôle visuel industriel	13
1.2.2	Formaliser le contrôle visuel humain	15
1.2.2.1	Décrire les anomalies	15
1.2.2.2	Modéliser les procédures de contrôle	16
1.2.2.3	Méthodologie d'exploration dans le cadre du contrôle	17
1.2.3	Vers une automatisation du contrôle visuel	19
1.2.4	Synthèse quant au contrôle visuel humain	23
1.3	État des lieux du contrôle automatique par vision industrielle	24
1.3.1	Automatisation et systèmes de vision industrielle	24
1.3.1.1	Classifications des applications des systèmes de vision industrielle	24
1.3.1.2	Points clefs pour la conception d'un système de vision industrielle	29
1.3.2	Positionnement applicatif du contrôle automatisé de la qualité d'aspect	31
1.3.2.1	Secteurs d'activités	31
1.3.2.2	Moyens pour traiter l'information	34
1.3.3	Cahier des charges d'un système de contrôle reproduisant les caractéristiques spécifiques d'un contrôle humain	36
1.4	Conclusion	38

1.1 Introduction

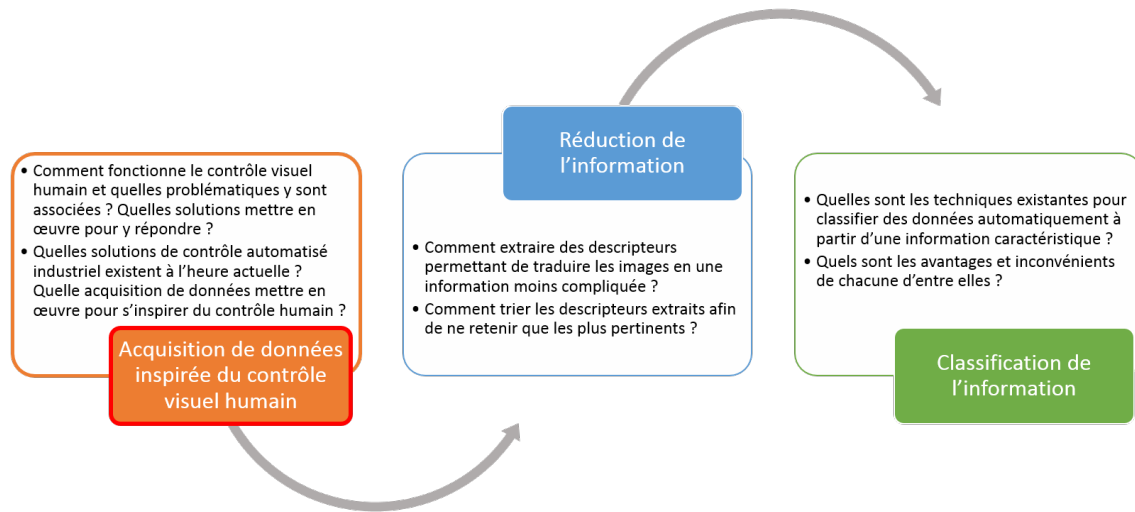


Figure 1.1 – Problématiques liées à l'acquisition de données inspirée du contrôle visuel humain

La perception humaine est au centre des processus de contrôle qualité dit sensoriels, comme le contrôle visuel d'aspect. Si elle est essentielle à prendre en compte pour juger de la *qualité perçue* d'un produit, elle entraîne également l'apparition d'une variabilité dans le contrôle humain.

La qualité sensorielle des produits est une problématique de recherche très active et prenant de plus en plus d'ampleur, au vu des intérêts industriels croissants. Différents travaux de thèses ont été menés depuis une dizaine d'années au laboratoire SYMME à ce sujet. Ils ont tout d'abord porté sur la modélisation du contrôle visuel et la réduction de la variabilité du contrôle humain avec notamment les thèses de Anne Sophie Guerra, Nathalie Baudet et Gaëtan le Goïc. Cette dernière thèse a également lancé l'évolution des travaux vers l'automatisation du contrôle, problématique portée ensuite par les thèses de Simon Desage et Gilles Pitard (cf. Avant Propos).

Ces présents travaux de thèse s'inscrivent pleinement dans la continuité des travaux déjà menés au laboratoire SYMME. Nous proposons donc de dédier en grande partie ce premier chapitre à un aperçu des résultats des travaux précédents, afin de permettre au lecteur de bien comprendre le point de départ et les orientations de nos propres travaux.

Ce chapitre s'applique tout d'abord à présenter rapidement les différentes sources d'influences qui peuvent entraîner la variabilité du contrôle visuel humain, ainsi que sa mise en évidence dans un contexte industriel.

Différents travaux ont ensuite porté sur la réduction de cette variabilité, en proposant des outils pour caractériser les anomalies d'aspect et formaliser les processus de contrôle visuel humain. Nous décrivons certains d'entre eux afin de mieux comprendre ce que sont les anomalies d'aspect et comment les contrôler.

Cette première partie cherche à répondre à la question suivante : "Comment fonctionne le contrôle visuel humain et quelles problématiques y sont associées ? Quelles solutions mettre en œuvre pour y répondre ?".

Enfin, nous justifions l'évolution de ces méthodes de contrôle vers des solutions automatisées dans

un contexte industriel. Un état des lieux de ces solutions est également développé afin d'identifier des pistes de développement vers un contrôle automatisé inspiré du contrôle humain.

Cette deuxième partie de chapitre cherche quant à elle à répondre aux questions suivantes : "Quelles solutions de contrôle automatisé industriel existent à l'heure actuelle ? Quelle acquisition de données mettre en œuvre pour s'inspirer du contrôle humain ?

Ce chapitre se conclura par une synthèse des éléments présentés, sous forme d'un cahier des charges pour guider la suite de notre travail.

1.2 Le contrôle visuel humain en industrie

Pour pouvoir garantir une qualité d'aspect satisfaisante, le contrôle visuel des produits tient une place essentielle dans les processus de contrôle qualité des produits sur les lignes de production. Ce type de contrôle fait parti de la catégorie des contrôles qualifiés de sensoriels, car faisant appel au sens visuel humain. La qualité visuelle des produits est donc contrôlée à travers le filtre de la perception humaine, ce qui introduit une certaine subjectivité dans l'inspection. Alors fondé sur des grandeurs non mesurables, le contrôle visuel est soumis à une variabilité de résultats. En effet, deux contrôleurs avec des perceptions différentes peuvent être amenés à juger la conformité d'une pièce de manière différente. Mais au delà d'une différence de perception entre deux personnes, le jugement d'un seul et même contrôleur peut également varier en fonction de son environnement, de son humeur, de son état de fatigue. Pour comprendre ces phénomènes, nous proposons tout d'abord de nous intéresser plus en détail à la perception humaine et à ces sources d'influence.

Nous souhaitons préciser ici que très peu de travaux universitaires publiés portent sur le contrôle visuel humain de produits industriels, en dehors de ceux menés précédemment au laboratoire SYMME. De ce fait, la bibliographie utilisée dans cette première partie se révèle très locale.

1.2.1 Perception humaine et variabilité du contrôle visuel

1.2.1.1 Perception et sources d'influence

La perception humaine peut être définie comme une "*opération psychologique complexe par laquelle l'esprit, en organisant les données sensorielles, se forme une représentation des objets extérieurs et prend connaissance du réel*"¹. De nombreuses théories ont été proposées, tout particulièrement en psychologie, pour décrire ce processus complexe. Si nous ne détaillons pas ces différentes théories dans ce mémoire, nous invitons le lecteur à consulter l'ouvrage de Bagot qui détaille ces différentes approches pour approfondir le sujet [Bagot1999]. De manière simplifiée, ce processus consiste en l'acquisition d'informations par nos sens, mais également en leur traitement pour leur donner du sens dans le monde qui nous entoure. Chacune de ces deux étapes peut être influencée à différents niveaux, entraînant une variabilité du jugement humain.

Pour commencer, l'acquisition des données visuelles par notre capteur oculaire, avant de débiter le processus de traitement des données, peut être influencée par divers facteurs. Dans le contexte de l'inspection visuelle en industrie, Debrosse propose un exemple d'étude d'influence dans laquelle il utilise la méthode des 5M de Ishikawa pour identifier ces facteurs (Figure 1.2) [Debrosse2010].

L'environnement de contrôle y tient une place importante, notamment en ce qui concerne les conditions d'éclairage. Des conditions d'éclairage différentes peuvent en effet faire varier l'image acquise par nos yeux. Une différence observée peut alors être due à une variation de l'éclairage et non à une réelle anomalie sur le produit.

Le produit en lui-même est aussi porteur de facteurs pouvant influencer la prise d'information. Sa taille ou encore les propriétés de sa surface (réflectivité, texture, etc.) peuvent par exemple induire des difficultés pour nos yeux à enregistrer une information pertinente et répétable.

Les méthodes utilisées pour le contrôle visuel peuvent aussi influencer l'acquisition d'information. En effet, si le processus d'exploration est trop aléatoire, l'information pertinente peut ne pas être captée, ou pas de manière répétable.

Enfin, le dernier grand poste d'influences réside dans le contrôleur lui-même. L'acuité visuelle et

1. Définition du Centre National de Ressources Textuelles et Lexicales (CNRTL)

le niveau expérience de la personne effectuant le contrôle sont des premiers facteurs pouvant influencer la qualité et la répétabilité de l'acquisition des informations. Mais même si le sujet n'a pas de problème d'acuité visuelle et est très expérimenté, d'autres facteurs comme son état de fatigue ou sa concentration peuvent rapidement faire également varier son jugement. En effet, certains travaux de recherche comme ceux présentés par Rensink ont montré que certaines modifications rapides et localisées dans une scène visuelle peuvent ne pas être perçues si l'attention du sujet est focalisée sur un autre événement visuel [Rensink1997].

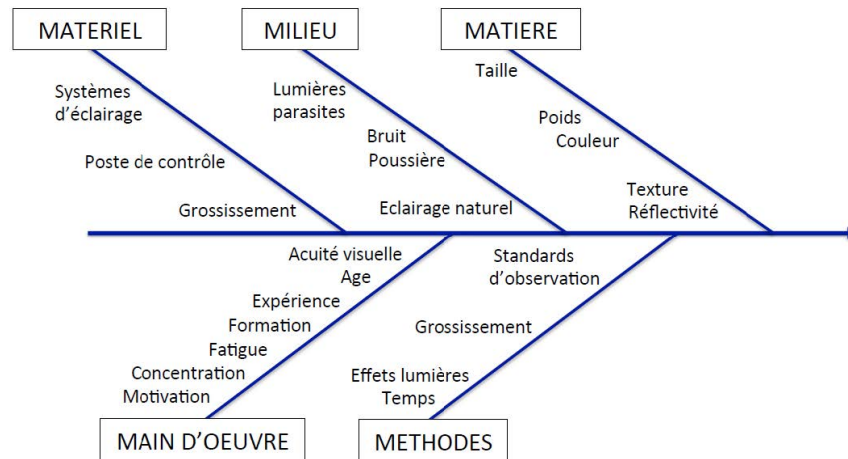


Figure 1.2 – Illustration de l'étude d'influences lors du contrôle visuel de surfaces dans un milieu industriel [Debrosse2010]

Ce dernier point fait le lien avec la deuxième étape de la perception : le traitement des informations acquises. L'humain est en effet soumis à divers processus mentaux qui résultent en un traitement subjectif des informations reçues par ses sens. De nombreuses théories ont été élaborées pour tenter de décrire au mieux les mécanismes mis en jeu, principalement d'un point de vue psychologique. Même si elles ne s'accordent pas toutes sur l'origine ou le fonctionnement détaillé de la perception, il peut en être synthétisé que l'interprétation des données sensorielles est influencée inconsciemment notamment par les expériences passées, les motivations, les besoins ou encore les valeurs d'un individu [Baudet2012].

De plus, des problématiques plus particulièrement liées à la perception visuelle apparaissent également lorsqu'il s'agit de juger une information sensorielle issue des yeux. En effet, la perception de formes, d'objets et de profondeur naît d'une combinaison de contrastes de lumière, de couleurs ou encore d'autres formes perçues dans l'environnement d'observation. Différentes approches comme la théorie de Gestalt, introduite par Wertheimer [Wertheimer1923], ou encore les approches de Treisman, de Biederman ou encore de Marr permettent notamment de mieux comprendre le fonctionnement de la perception des formes et des objets. Nous ne détaillons pas ces théories dans ce manuscrit, mais nous invitons le lecteur à consulter les livres de Bagot et Weil-Barais [Bagot1999] [Weil-Barrais1996] qui traitent plus longuement de ces approches. L'organisation des contrastes, des couleurs et des formes environnants à l'objet observé peuvent donc également influencer la perception finale d'un individu. Ces phénomènes peuvent être mis en évidence grâce à des illusions d'optique, dans lesquelles le traitement de l'information visuelle dans son environnement nous dupe dans l'interprétation réelle d'une image.

L'échiquier d'Adelson, proposé en 1995 par Adelson, est un exemple d'illusion portant sur la perception du contraste et de la couleur d'une image [Adelson2005]. Le problème y est de déterminer les couleurs des deux cases A et B. Généralement, celles-ci sont interprétées comme

différentes par le cerveau alors qu'en réalité elles ont la même teinte, comme illustré en Figure 1.3. L'illusion est créée grâce à l'association de plusieurs facteurs.

Le premier repose sur le fait que chacune des deux cases est entourée par des cases de teintes plus foncées ou plus claires, ce qui renforce le contraste dans un sens ou dans l'autre.

Ensuite, la présence de l'ombre portée issue du cylindre vert joue également un rôle important. En effet, le cerveau attribue une clarté supérieure à la case B car elle se situe dans l'ombre. Il a tendance à compenser les effets des ombres afin de pouvoir déterminer la couleur d'une même surface sous différentes conditions d'éclairages.

Enfin, un dernier facteur apparaît également dans l'utilisation d'un échiquier, objet familier pour la plupart des observateurs. La connaissance à priori de cet objet introduit une interprétation visant à se rapprocher de l'expérience plutôt que du stimulus réel reçu par l'œil.

Cette illusion montre donc la capacité de la perception humaine à interpréter le contraste et les couleurs d'une image pour donner du sens aux objets qui la compose. Mais elle révèle également un risque de déformation de l'information réelle.

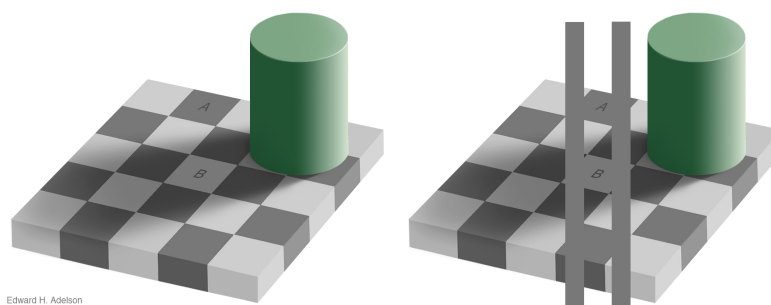


Figure 1.3 – L'échiquier d'Adelson initial et révélé

Dans un autre registre, d'autres illusions mettent en évidence les processus mis en place par le cerveau humain pour reconnaître des formes. Certaines des plus connues reposent sur la théorie de Gestalt, ou théorie des formes, qui postule que l'humain perçoit son environnement comme un ensemble d'éléments structurés selon un ensemble de principes introduits par Wertheimer [Wertheimer1923] :

- **Le principe de proximité** : des éléments proches les uns des autres sont perçus comme associés les uns aux autres, par exemple comme appartenant à un groupe, ou à un plus grand ensemble.
- **Le principe de similarité** : des éléments partageant des propriétés communes tendent à se regrouper les uns avec les autres.
- **La loi de destin commun** : des éléments en mouvement partageant une même trajectoire sont perçus comme appartenant au même ensemble.
- **Le principe de la bonne forme**, ou loi de la simplicité ou loi de la prégnance : lorsque la perception humaine est confrontée à un ensemble d'éléments, elle tend à percevoir la plus simple, la plus équilibrée, la plus stable configuration en premier.
- **Le principe de continuité**, ou de bonne continuité : si des éléments rapprochés sont perçus comme formant une ligne qui apparaît comme ayant une ou plusieurs branches, la branche qui suit la direction du trait d'origine est perçue comme étant la continuité du parcours, et les autres sont perçues comme annexes.
- **Le principe de clôture** : la perception humaine a tendance à relier les éléments afin qu'ils constituent des formes fermées plutôt qu'ouvertes.
- **Le principe de familiarité** : le cerveau a tendance à reconnaître des formes familières en priorité.

Ces principes sont illustrés par des illusions d'optiques en Figure 1.4, et mettent en évidence la capacité de la perception humaine de reconnaître des formes à partir d'éléments visuels simples. De même que pour l'exemple précédent de l'échiquier, cette interprétation mise en place par le cerveau est essentielle à appréhender le monde qui nous entoure, mais peut également nous duper sur l'information réellement reçue.

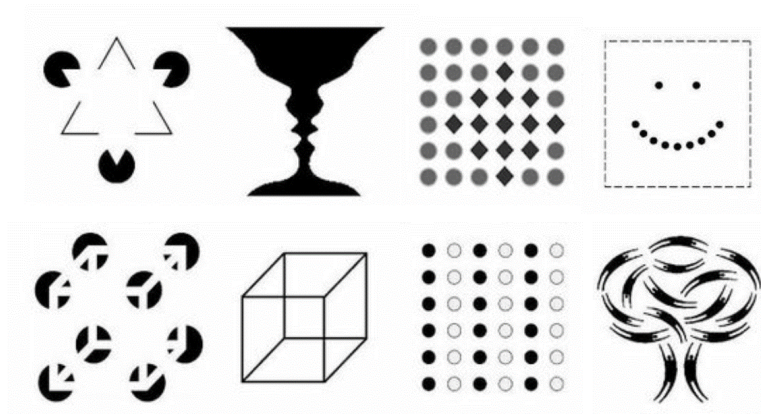


Figure 1.4 – Illusion d'optiques inspirées des lois de Gestalt

Tout cela induit donc un risque individuel de défaillance de la perception, l'information réelle pouvant être déformée au cours de son interprétation. La taille, la forme ou encore la couleur d'un objet dans une image peuvent alors être mal jugées par l'observateur.

De plus, les acuités visuelles mais aussi les expériences, les attentes ou encore les valeurs de chacun étant différentes, la perception d'une même scène peut être divergente d'un individu à un autre. Il est même rare de trouver deux personnes ayant une perception totalement identique pour une même observation.

Si la perception humaine est un outil puissant pour analyser et comprendre le monde qui nous entoure, les problématiques relevées précédemment montrent également qu'elle peut être trompée sous certaines circonstances. Dans un contexte industriel, cela peut entraîner des erreurs de jugement lors du contrôle, qui se traduisent par une variabilité dans les résultats du contrôle humain.

1.2.1.2 Variabilité des résultats du contrôle visuel industriel

Dans le contexte du contrôle visuel industriel, la perception humaine est donc à la fois essentielle et problématique. Essentielle dans un premier temps car c'est bien la qualité perçue par le futur client qui doit être évaluée. Le filtre de la perception humaine est donc fondamental pour traduire cette qualité. Problématique ensuite car elle induit une variabilité des résultats du contrôle, liée aux sources d'influences présentées précédemment.

Pour mesurer et mettre en évidence cette variabilité, il est possible d'utiliser une analyse de variabilité, comme le test R&R (Reproductibilité et Répétabilité) par exemple. Ce type d'analyses permet de détecter des problèmes de répétabilité (résultats du contrôle différents pour un même produit et un même contrôleur à différents instants) et de reproductibilité (résultats du contrôle différents pour un même produit et plusieurs contrôleurs différents), comme illustré en Figure 1.5. Dans le cadre du contrôle visuel, le test R2&E2 (Reproductibilité et Répétabilité pour l'Exploration et l'Evaluation) est plus approprié car il permet d'être plus précis quant à l'origine de la variabilité [Maire2013b]. En effet, ce test analyse séparément la phase d'exploration et d'éva-

luation des surfaces lors du contrôle, et permet d'étudier la répétabilité et la reproductibilité des contrôleurs lors de ces deux phases.

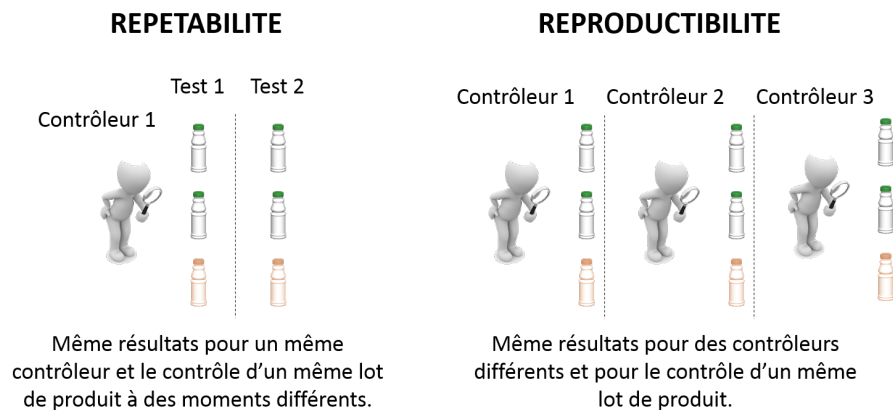


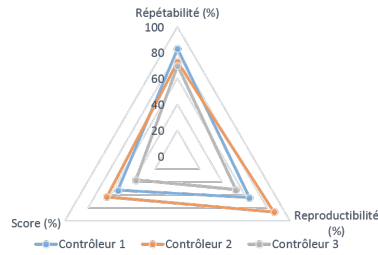
Figure 1.5 – Illustration de la répétabilité et reproductibilité dans le contexte du contrôle visuel

Baudet donne plusieurs exemples d'utilisation de cette analyse sur des cas industriels dans les entreprises Fournier et Eudica. Les résultats, résumés en Figure 1.6, permettent de se rendre compte des différents types de variabilités qui peuvent apparaître [Baudet2012]. Pour chaque cas d'étude, 6 indicateurs sont mesurés pour 3 contrôleurs :

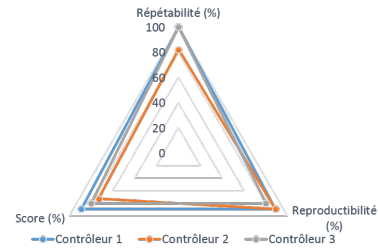
- La répétabilité du contrôleur en exploration, qui évalue sa capacité à *détecter* la même anomalie lors de deux contrôles d'un même produit.
- La reproductibilité du contrôleur en exploration, qui évalue sa capacité à *détecter* la même anomalie qu'un groupe d'experts, dans le cas où la répétabilité est bonne sur les deux contrôles.
- La répétabilité du contrôleur en évaluation, qui évalue sa capacité à *juger* une anomalie de la même manière lors de deux contrôles d'un même produit.
- La reproductibilité du contrôleur en évaluation, qui évalue sa capacité à *juger* une anomalie de la même manière qu'un groupe d'experts, également dans le cas où la répétabilité est bonne sur les deux contrôles.
- Les scores du contrôleur dans les deux cas, qui évalue la capacité du contrôleur à être à la fois répétable et reproductible sur l'ensemble des anomalies des tests.

Le premier constat porte sur la répétabilité des résultats du contrôle insuffisante chez certains contrôleurs, notamment dans le cas d'étude chez Fournier pour l'exploration. Ceci illustre bien une *variabilité intra-contrôleur* lors du contrôle. Un contrôleur peut ainsi détecter ou juger des anomalies de manières différentes lors de deux contrôles pour un même produit.

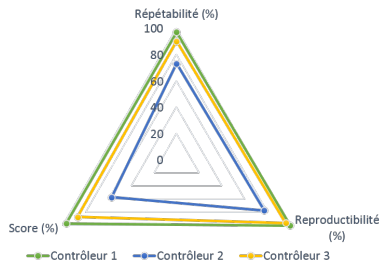
Ensuite, l'autre constat important réside dans les différences de performances inter-contrôleurs. Le cas d'étude chez Eudica permet particulièrement d'illustrer cela avec des performances bien inférieures pour le contrôleur 2. Il s'est avéré dans ce cas que le contrôleur 2 était novice par rapport aux deux autres, et qu'il ne disposait pas des informations et outils nécessaires pour effectuer son travail efficacement, particulièrement pour la phase d'évaluation. De manière plus générale, les différences de performances constatées mettent en lumière une *variabilité inter-contrôleur* qui dérive naturellement des différences de perception d'un individu à un autre. Cette variabilité s'illustre également par une reproductibilité insuffisante chez certains contrôleurs, synonyme d'une mauvaise capacité à détecter et/ou juger les anomalies de la même manière que le groupe d'expert.



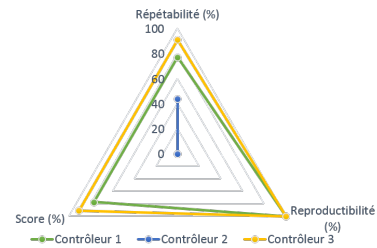
(a) R2&E2 Fournier Exploration



(b) R2&E2 Fournier Evaluation



(c) R2&E2 Eudica Exploration



(d) R2&E2 Eudica Evaluation

Figure 1.6 – Résultats de tests R2&E2 mené par Baudet chez Fournier et Eudica [Baudet2012]

Ces tests menés en industrie illustrent donc l'existence d'une variabilité des résultats du contrôle à la fois *intra-contrôleur* et *inter-contrôleurs*. Comme le souligne Baudet, cela justifie les difficultés qu'ont les entreprises à mettre en place un contrôle visuel fiable [Baudet2012]. Mesurer et localiser cette variabilité lors des différentes étapes du contrôle, notamment avec le test R2&E2, est déjà un premier pas vers la réduction de la variabilité, car elle peut permettre de comprendre plus précisément ses origines. Pour aller plus loin dans la réduction de cette variabilité, des travaux ont également porté sur la modélisation et la formalisation du contrôle, afin de proposer des procédures standardisées aux contrôleurs.

1.2.2 Formaliser le contrôle visuel humain

La variabilité des résultats du contrôle visuel peut-être tout d'abord expliquée par la diversité des pratiques existantes, liée à un manque de spécifications et de références dans le domaine du contrôle visuel [Debrosse2010]. Une première action corrective apparaît alors dans la proposition de standards pour caractériser les anomalies visuelles, modéliser les procédures de contrôle et proposer des méthodologies d'inspection génériques [Maire2013a] [Baudet2011] [Baudet2012] [Guerra2008]. En effet, si tout le monde contrôle de la même manière, cela permet déjà de réduire la subjectivité du contrôle, et donc sa variabilité.

1.2.2.1 Décrire les anomalies

Dans l'objectif de formaliser le contrôle, la description des anomalies visuelles constitue une première étape. Guerra et Baudet se sont intéressées à la conceptualisation du contrôle humain, en commençant par des études des pratiques dans le milieu industriel. Un premier constat fait par Guerra a été le foisonnement de descripteurs utilisés pour décrire les anomalies lors des contrôles visuels, n'aidant pas les contrôleurs à prendre une décision répétable. Grâce à une méthodologie basée sur l'analyse sensorielle, méthode permettant de réaliser l'étude subjective d'un produit, elle a proposé une réduction à 4 descripteurs pour classer toute anomalie rencontrée, dans le contexte

de l'entreprise Patek Philippe SA [Guerra2008]. Baudet a repris cette classification et l'a adaptée pour qu'elle soit applicable dans tous types d'entreprises et pour tous types de produits (Tableau 1.1) [Baudet2012].

Descripteur	Définition
Marque	Ce qui abîme la surface, qui génère une rupture dans la forme (ex. : rayure, piqure, coup, ...)
Hétérogénéité	Ce qui fait perdre à la surface son homogénéité (ex. : tâche, différence de couleur, de texture, ...)
Pollution	Ce qui "pollue" la surface du produit (ex. : poil, poussière, point noir, ...)
Déformation	Ce qui modifie la forme de la surface du produit (ex. : ligne de lumière irrégulière, "trop arrondi", ...)

Tableau 1.1 – Les 4 descripteurs et leur définition

Cette diminution de la complexité du vocabulaire et des descripteurs utilisés pour caractériser les anomalies visuelles a été un premier pas pour modéliser et formaliser le contrôle.

1.2.2.2 Modéliser les procédures de contrôle

Au delà de la description des anomalies, la modélisation des procédures pour l'exercice du contrôle permet également de guider les contrôleurs dans leur prise de décision quant à la conformité des pièces. Baudet a proposé de décomposer le contrôle en trois étapes distinctes [Baudet2012] (Figure 1.7), dérivant d'un modèle proposé initialement par Guerra [Guerra2008] :

- L'**exploration** consiste en l'exploration de la surface et éventuellement la détection d'un écart, appelé anomalie. Baudet catégorise tout d'abord les anomalies en deux grands types : évolutive et non-évolutive. Dans le cas où l'anomalie est évolutive, elle est directement rejetée. Dans le cas où elle est non-évolutive, elle est caractérisée suivant les catégories définies précédemment (Tableau 1.1). Si aucune anomalie n'est observée, la pièce peut être directement déclarée comme conforme.
- L'**évaluation** comporte une phase caractérisation de l'anomalie par des attributs sensoriels définis en fonction du type d'écart détecté puis d'une phase de quantification de son intensité. Baudet définit la combinaison de ces attributs et de l'intensité comme le *profil sensoriel* de l'anomalie. Il est à noter que pour Baudet, la phase de localisation de l'anomalie fait partie de l'évaluation de son profil sensoriel, alors qu'il faisait parti de la phase de décision chez Guerra. En effet, le contexte de localisation de l'anomalie peut influencer sa caractérisation et doit donc faire partie de la phase d'évaluation.
- La **décision** consiste en la localisation et spécification de l'anomalie, c'est-à-dire la prise de décision quant à la conformité du produit. Cette décision doit prendre en compte la gamme du produit ou encore le client, facteurs qui peuvent faire varier le seuil entre conformité et non-conformité.

Ces principes ont ensuite été plus amplement détaillés et développés en termes de méthodes et d'outils dans ces deux thèses, et ont également été validés dans différents contextes et entreprises industriels de la région annécienne et genevoise (Patek Philippe, Fournier, S.T. Dupont, NTN SNR et Eudica). Les méthodologies issues de ces thèses sont d'ailleurs maintenant utilisées dans beaucoup d'entreprises de ce bassin d'activité.

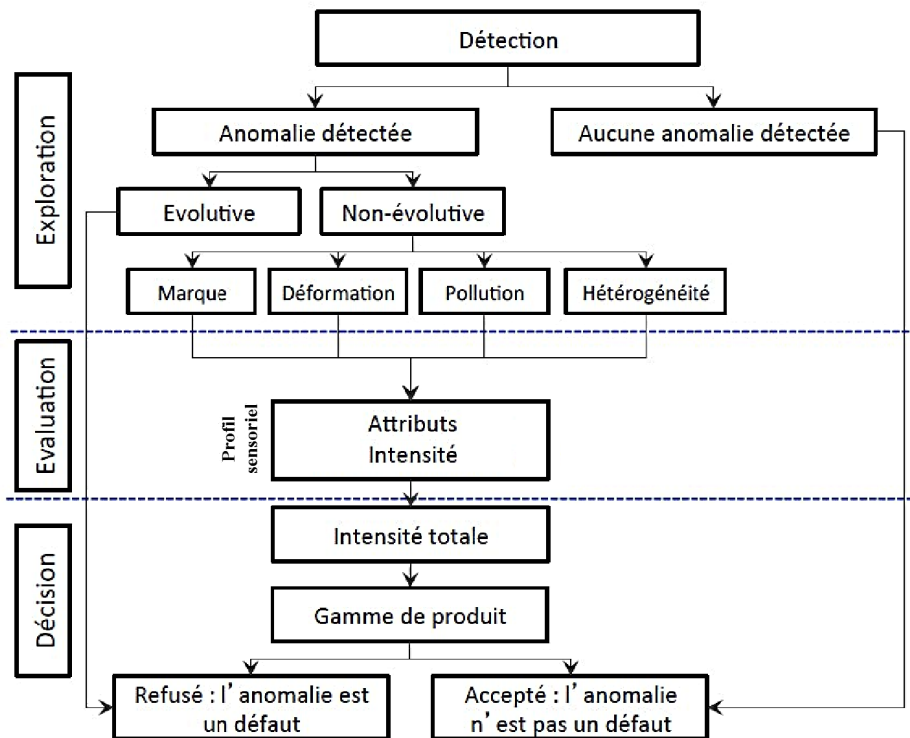


Figure 1.7 – Modèle pour le contrôle sensoriel proposé par Baudet [Baudet2012]

1.2.2.3 Méthodologie d'exploration dans le cadre du contrôle

Suite au développement du modèle précédent, des méthodes et recommandations peuvent être proposées pour chaque phase du contrôle. Nous avons décidé de présenter un peu plus en détail ici les recommandations et pratiques mises en œuvre pour la phase d'exploration, et plus particulièrement en terme de méthodes d'observation. En effet, certaines des notions abordées permettront de mieux comprendre certains choix effectués par la suite dans ce travail de thèse.

Lors de la phase d'exploration, les contrôleurs sont amenés à se placer dans différentes configurations d'observation, appelée *scènes d'observation*. Ils sont alors amenés à faire varier l'angle de la lumière incidente et l'angle d'observation par rapport à la normale à la surface de l'objet, pour faciliter la détection d'anomalies. Dans ce cadre, 3 configurations particulières, nommées *Effets lumière*, permettent souvent de mettre en valeur les anomalies visuelles. Il s'agit de l'effet lumière noire, de l'effet brillance ou lumière blanche, et de l'effet miroir. Les caractéristiques de chaque effet sont développées en Figure 1.8. Dans la pratique, le contrôleur humain fait "miroiter" la pièce, ce qui consiste à la faire tourner dans sa main afin de balayer successivement différentes configurations angulaires dans le but de rechercher les effets lumière.

Partant de ces considérations, il est possible de proposer des méthodologies d'exploration spécifiques à l'inspection visuelle d'un produit, afin d'aider les contrôleurs à détecter les anomalies. Le Goïc nous fournit un exemple avec une gamme d'exploration proposée dans le cadre du contrôle de briquet ST Dupont (Figure 1.9). Il propose ainsi différents mouvements angulaires à effectuer pour les différentes faces à inspecter, mais aussi un parcours des yeux particuliers. Une telle gamme ne peut être construite qu'à partir d'une connaissance des anomalies susceptibles d'apparaître sur le produit et est donc le résultat d'un travail amont de recensement et de caractérisation des anomalies.

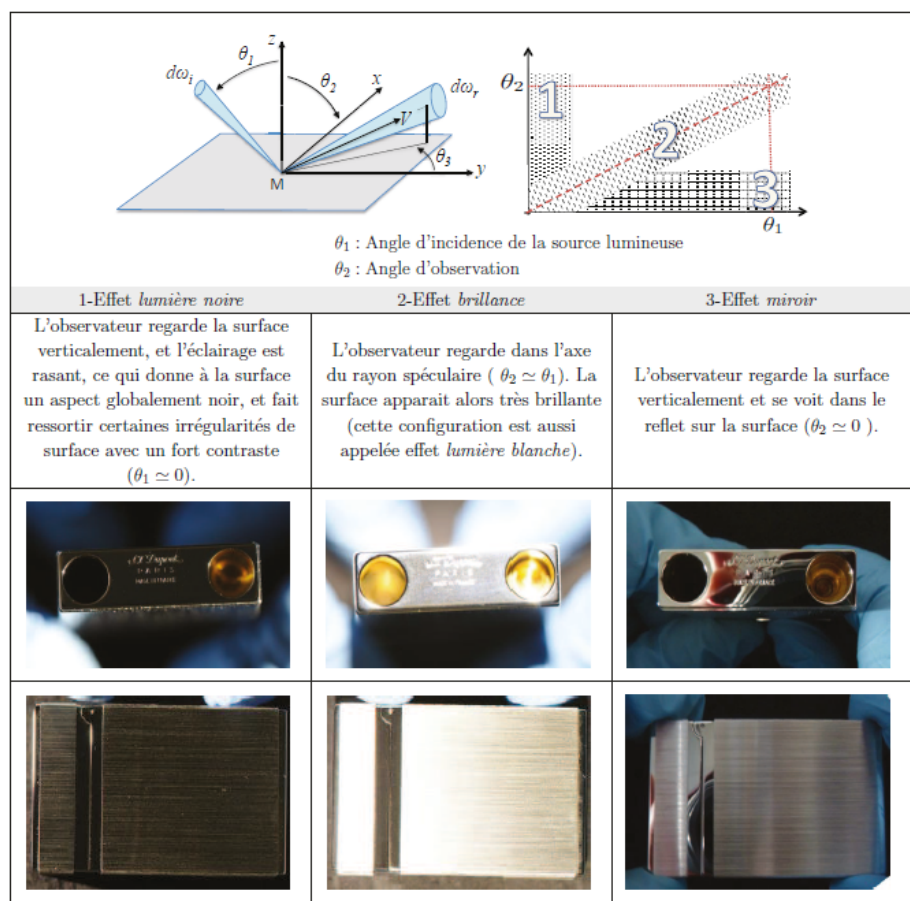


Figure 1.8 – Les 3 effets lumière et leur caractéristiques [Le Goïc2012]

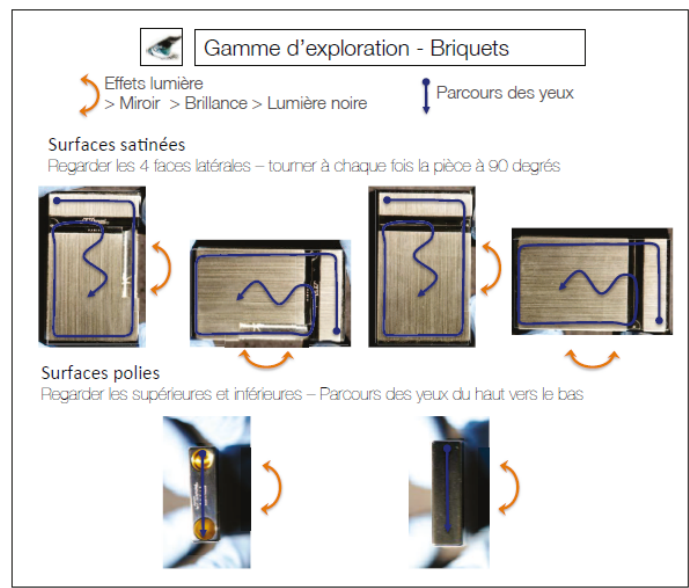


Figure 1.9 – Exemple de gamme d'exploration proposée par [Le Goïc2012]

En conclusion, afin de pouvoir détecter différentes anomalies si elles sont présentes, le contrôleur est amené à observer la pièce suivant différents points de vue et en jouant avec la lumière incidente pour rechercher certains effets appelés effets lumières. Cela met en lumière le fait que certaines anomalies ne sont visibles que suivant certains cônes de vision limités [Le Goïc2012]. La

compréhension de ce phénomène permet de fournir des parcours d'exploration standardisés aux contrôleurs, adaptés à chaque produit. Ceux-ci permettent aux contrôleurs de parcourir un grand nombre de scènes d'observations, ce qui facilite la détection des anomalies.

1.2.3 Vers une automatisation du contrôle visuel

Les travaux précédents visent donc à réduire la variabilité de l'inspection humaine en proposant des outils de formalisation et de standardisation des procédures de contrôle. Mais avec le développement récent des technologies numériques et l'amélioration importante de leurs performances en terme de calcul, de nombreux développements tentent également de reproduire le contrôle via des systèmes automatisés. Il s'agit alors d'allier la performance de l'humain à prendre des décisions compliquées via son esprit rationnel et critique, à la puissance de calcul de machines présentant une répétabilité sans failles. Nous présentons ici certains principes pour cette automatisation de l'inspection d'aspect, au regard du modèle du contrôle humain.

Dans l'optique de faire le parallèle entre l'inspection visuelle humaine et la vision industrielle réalisée par des machines, Desage propose d'adapter le modèle proposé par Baudet, en l'orientant vers une application automatisée du contrôle (Figure 1.10) [Desage2015]. Pour aller dans ce sens, il propose une réorganisation des 3 phases précédentes, tout en identifiant différentes étapes du contrôle humain avec des éléments renvoyant à une stratégie de vision industrielle :

- **L'exploration** : Elle se limite maintenant uniquement à la *Détection* d'un ou des écarts sur la surface par rapport à un modèle virtuel (CAO ou intention du constructeur), à un modèle réel (pièce témoin) ou au voisinage local (rupture de continuité). Comme chez Baudet, la non détection d'anomalie renvoie directement à l'acceptation du produit. Par contre, pour Desage, la phase d'exploration s'arrête avec l'étape de détection. La caractérisation de l'anomalie est alors déplacée dans la phase d'évaluation.
- **L'évaluation** : Elle se compose d'une étape de *Reconnaissance*, qui reprend la caractérisation de l'anomalie proposée par Baudet, et d'une étape d'*Évaluation* permettant d'évaluer l'anomalie suivant différents niveaux d'échelles.
- **La décision** : Elle permet de donner une *Conclusion* quant à la qualité finale du produit : conforme (accepté) ou non conforme (refusé).

Pour faire le parallèle avec les systèmes de vision par ordinateur, Desage identifie tout d'abord l'étape d'observation avec celle d'obtention d'une image numérique. Ensuite, les 4 étapes de Détection, Reconnaissance, Évaluation et Conclusion sont identifiées comme 4 cycles répétés d'extraction d'attributs suivie d'une classification de ceux-ci. Les 3 premiers cycles servent à identifier les attributs pertinents pour évaluer le produit, et le dernier à obtenir une classification binaire des produits à partir de ces attributs (Figure 1.11).

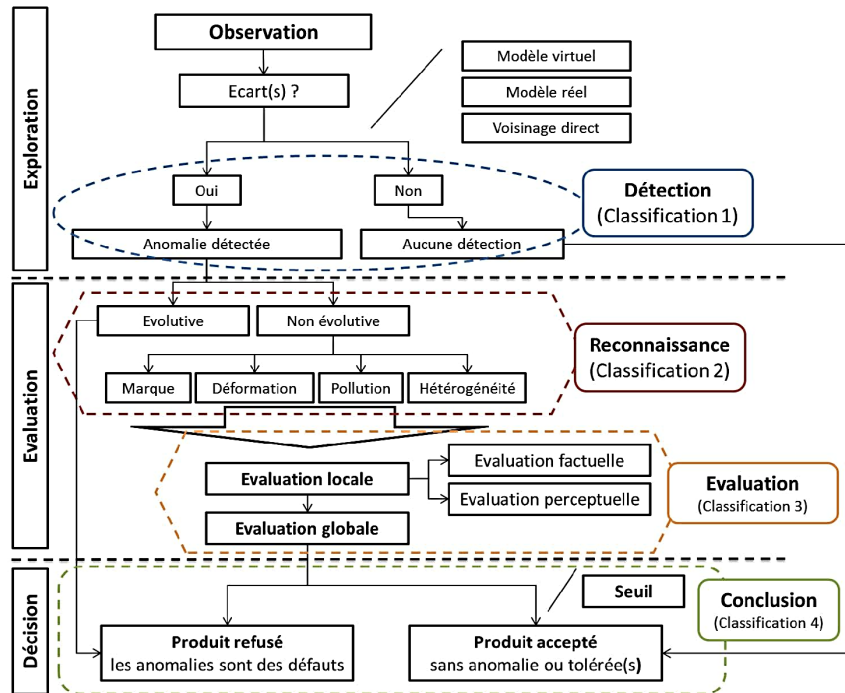


Figure 1.10 – Modèle pour le contrôle sensoriel au regard de l'automatisation proposé par Desage [Desage2015]

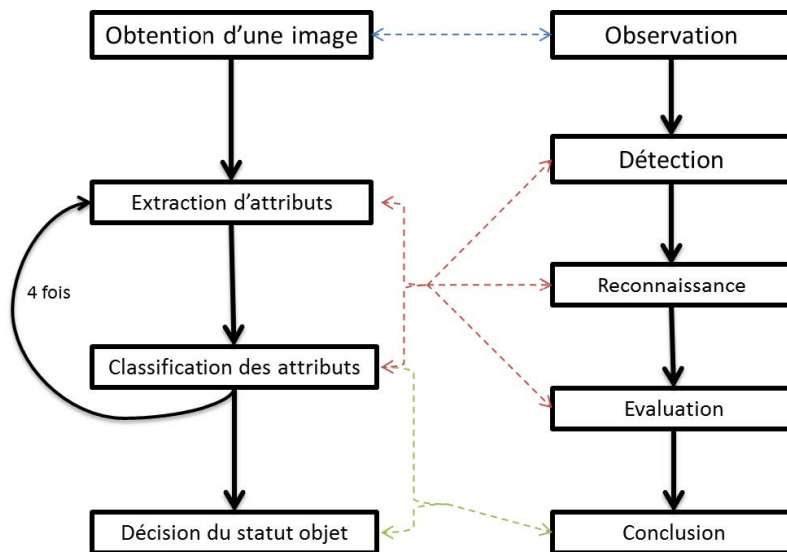


Figure 1.11 – Parallèle entre le modèle du contrôle sensoriel et une stratégie de vision industrielle selon Desage [Desage2015]

En pratique, les stratégies de vision industrielle comportent généralement un seul cycle d'extraction d'attributs et de classification qui permet d'aller directement à la conclusion finale. Nous proposons donc d'adapter l'interprétation proposée par Desage, pour la rendre plus proche des applications réelles. Pour cela, il est tout d'abord nécessaire de détailler un peu plus la composition d'un système typique de vision par ordinateur (Figure 1.12). Celui-ci a été décrit dans diverses publications [Awcock1995] [Malamas2003] [Golnabi2007].

Celui-ci commence avec l'acquisition d'une image numérique depuis une scène éclairée par une source de lumière. Cette image numérique est obtenue grâce à un capteur optique (par exemple un

capteur CCD ou CMOS) qui transforme l'image optique en signal électrique puis enfin en image numérique.

S'en suit une étape de prétraitement permettant d'obtenir une image numérique la mieux exploitable possible pour la suite. Parmi les opérations de prétraitement les plus utilisées, on retrouve les modifications de contraste ou de luminosité, l'égalisation d'histogramme, le lissage ou encore la rotation.

L'image est généralement ensuite segmentée afin d'identifier des régions d'intérêt qui correspondent à des parties de l'objet sur lesquelles les traitements suivants seront exécutés.

L'extraction d'attributs correspond à l'identification de paramètres permettant de réduire l'information de l'image de manière pertinente. Ces attributs deviennent des caractéristiques de l'objet. Enfin, l'étape de classification permet de classer l'objet en fonction des valeurs des attributs extraits. La classification peut-être binaire comme dans le cas d'une inspection de conformité dans les modèles précédents, mais elle peut aussi se fonder sur un plus grand nombre de classes, si l'on souhaite par exemple que le système puisse prédire quel type d'anomalie est présent sur la surface. Le résultat de la classification entraîne ensuite des actions sur la scène, comme la mise au rebut de la pièce si elle est jugée non conforme.

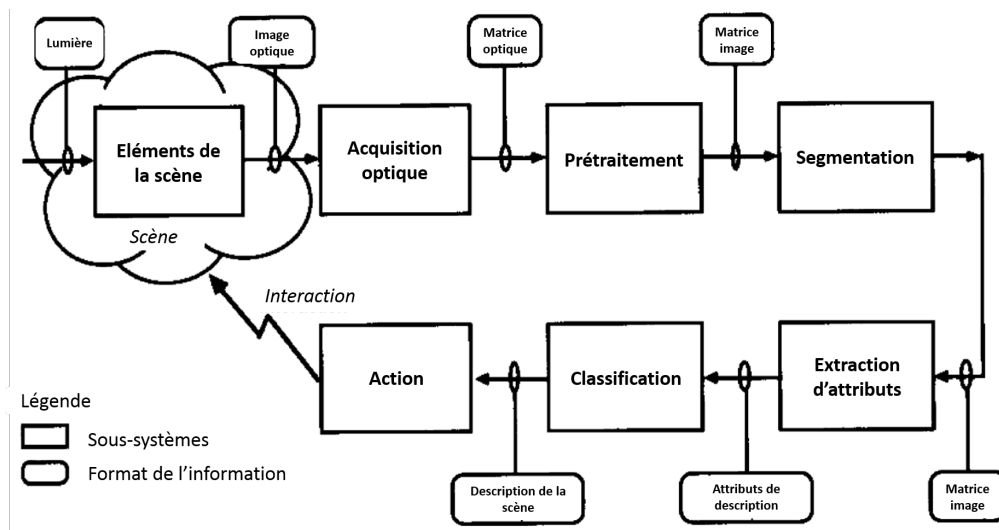


Figure 1.12 – Schéma bloc d'un système typique de vision par ordinateur [Awcock1995]

A la lumière de ce modèle, il apparaît donc bien qu'un système de vision par ordinateur n'exécute donc qu'un seul cycle extraction d'attribut/classification, et nous proposons donc un nouveau modèle faisant le parallèle entre le processus automatisé et l'inspection humaine, inspiré de celui présenté par Desage (Figure 1.13). De plus, nous proposons également de détailler un peu plus la phase d'obtention d'une image.

Chez Desage, la phase d'observation est mise en parallèle de l'obtention d'une image dans la stratégie de vision par ordinateur. Nous proposons ici de détailler l'obtention de l'image en 3 étapes, faisant parties du modèle d'un système typique de vision par ordinateur :

- L'acquisition optique : elle correspond à la traduction de l'image optique en une matrice numérique, ou image numérique, par un capteur optique.
- Le prétraitement : il s'agit d'une suite d'opérations permettant de modifier et de préparer les pixels pour extraire au mieux l'information de l'image numérique par la suite.
- La segmentation : elle consiste en un découpage de l'image en régions d'intérêt, communément appelées ROI (Region Of Interest en anglais). Ces régions sont identifiées comme contenant l'information pertinente de la surface observée. En effet, l'analyse de la scène

complète est rarement pertinente en pratique.

Pour le contrôle humain, la phase d'observation consiste en l'exploration visuelle de la surface. Il est intéressant de remarquer que lors de cette phase, le contrôleur est souvent amené à se focaliser sur des zones spécifiques identifiées par avance comme étant les plus critiques en terme de présence d'anomalies. Cette focalisation correspond à l'étape de segmentation dans le processus automatisé.

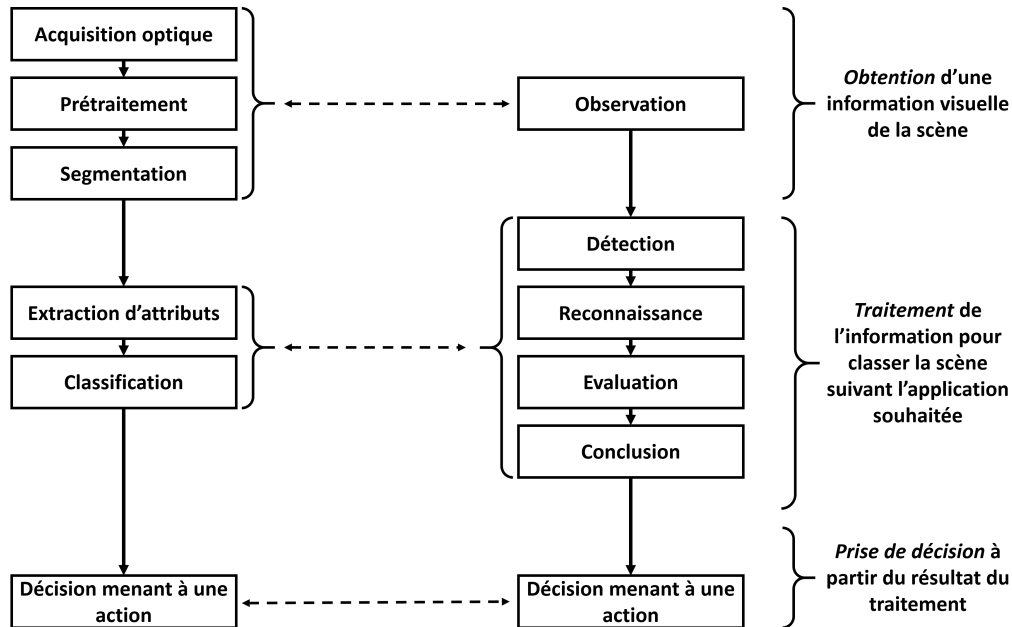


Figure 1.13 – Parallèle entre le modèle de contrôle sensoriel et une stratégie de vision industrielle plus détaillée

Comme le précise Malamas, la décision donnée par le système automatisé dépend en réalité de l'application, donc de l'objectif pour lequel il est développé [Malamas2003]. Il serait donc envisageable de développer un système exécutant 4 fois le cycle d'extraction et classification pour donner une décision lors de chaque étape de détection, reconnaissance, évaluation et conclusion, comme le propose Désage. Or, dans le contexte de l'inspection industrielle, la phase de classification donne généralement directement une conclusion finale sur la qualité du produit, prenant en compte à la fois la détection ou non d'une anomalie, et si c'est le cas sa reconnaissance et son évaluation. Les 4 étapes présentes dans le contrôle humain sont dans ce cas réalisées en même temps par le système automatisé.

A la suite de la phase d'observation, nous proposons donc de mettre en parallèle les 4 étapes de détection, reconnaissance, évaluation et conclusion, avec une seule étape regroupant une extraction d'attributs et une classification, et non plus 4 répétitions de ce cycle.

Enfin, les étapes de classification et de conclusion mènent toutes les deux à une décision quant à la qualité du produit, et entraînent une action, qui peut être par exemple sa mise au rebut si elle n'est pas conforme.

Au final, il est possible de synthétiser ce parallèle en trois grandes phases :

- L'obtention d'une information visuelle de la scène d'observation, adaptée à l'application souhaitée.
- Le traitement de cette information dans le but de classer la scène observée. La classification est définie par l'application souhaitée. Elle peut par exemple se résumer en une classification conforme/non conforme.

- La prise d'une décision consécutive au résultat du traitement, menant à une action sur la scène d'observation.

1.2.4 Synthèse quant au contrôle visuel humain

Les travaux présentés jusqu'ici se sont focalisés sur l'étude du contrôle visuel humain, et ont permis de mettre en évidence la présence d'une variabilité inter et intra-contrôleurs lors de son exercice.

La variabilité intra-contrôleur peut s'expliquer par le fait que notre perception individuelle est influencée par de nombreux facteurs qui peuvent affecter la répétabilité du contrôle.

La variabilité inter-contrôleur peut s'expliquer par le fait que chaque individu possède par exemple une acuité visuelle ou des expériences différentes, ce qui induit une part de subjectivité qui altère la reproductibilité du contrôle.

Afin de réduire cette variabilité, une formalisation du contrôle est proposée, basée sur différentes préconisations :

- La description plus efficace des anomalies en réduisant le vocabulaire utilisé.
- Une modélisation générale du contrôle en le décomposant en différentes étapes : exploration, évaluation et conclusion.
- La proposition de méthodologies standardisées pour chaque étape, permettant aux contrôleurs de contrôler de la même manière. En ce qui concerne plus particulièrement la phase d'exploration, des gammes sont proposées pour indiquer aux contrôleurs quels sont les parcours visuels à effectuer et quels angles d'observation à balayer.

Enfin, la mise en parallèle de modélisation du contrôle humain avec une stratégie de contrôle automatisé permet d'identifier les différentes étapes entre-elles dans l'objectif de développer des systèmes automatisés au regard des pratiques humaines.

1.3 État des lieux du contrôle automatique par vision industrielle

La deuxième partie de ce chapitre porte sur un état des lieux des systèmes de vision industrielle, et plus particulièrement sur ceux dédiés au contrôle d'aspect.

En effet, les systèmes de vision par ordinateur présentent plusieurs applications possibles dans le milieu industriel. Un rapide aperçu de ces dernières est tout d'abord présenté par la suite, avant de se focaliser sur les systèmes dédiés au contrôle automatique d'aspect.

Dans ce cadre, nous abordons ensuite la question des différents secteurs industriels d'application, afin de justifier l'effort de thèse dans le domaine applicatif choisi.

Enfin, c'est l'état des recherches en termes de différentes techniques mises en œuvres dans ces systèmes qui est développé, afin de mieux appréhender les pistes explorées par la suite.

1.3.1 Automatisation et systèmes de vision industrielle

L'introduction de l'automatisation dans le milieu industriel a engendré une véritable révolution dans la gestion des procédés et processus de production. Les systèmes de vision y tiennent une place privilégiée car ils permettent de superviser différentes tâches automatisées, notamment sous des contraintes rendant la tâche difficile à l'être humain. C'est par exemple le cas lorsque les cadences de productions sont élevées, ou que l'environnement de production est hostile au travail humain. Au delà de leur rôle dans les tâches automatisées, ces systèmes favorisent également le développement d'outils de production plus flexibles, c'est à dire capables de s'adapter rapidement à des changements de produits ou encore de cadences [Kostal2010].

Cette première section s'applique à donner un aperçu des applications possibles de ce type de systèmes, ainsi que des points clefs à prendre en considération lors de leur conception.

1.3.1.1 Classifications des applications des systèmes de vision industrielle

Les systèmes de vision trouvent donc de nombreuses applications dans le milieu industriel. Le fonctionnement et la composition de tels systèmes a déjà été présenté dans la partie précédente, notamment avec la Figure 1.12, pour en faire le parallèle avec l'inspection visuelle humaine. Mais les applications de tels systèmes sont plus larges que les seules problématiques d'inspection automatisée. Afin de donner un aperçu de celles-ci, plusieurs classifications ont été proposées dans la littérature, suivant des approches différentes. Nous proposons d'en présenter trois ici, issues de [Malamas2003] et [Golnabi2007], afin de permettre au lecteur d'appréhender différents aspects de la mise en œuvre de ces dispositifs.

Classification axée sur l'objectif visé

Selon Golnabi, les différentes applications des systèmes de vision industrielle peuvent tout d'abord être regroupées en 4 catégories, fondées sur l'objectif visé lors de la mise en œuvre du système [Golnabi2007]. Cette première classification permet de donner un aperçu global de l'ensemble des applications possibles pour un système de vision :

- **L'inspection visuelle de produits** dont le rôle principal est de vérifier si les produits répondent bien à des spécifications données, et tient notamment un rôle important dans les contrôles qualité.
- **L'identification de produits** qui permet de faire la différence entre différentes pièces pour des objectifs d'assemblage, de transport ou encore de tri. Il est à noter que cette catégorie est très liée à la précédente, notamment pour les applications de tri dans le cadre du contrôle qualité.
- **Le contrôle de processus et procédés** qui a trait au diagnostic non plus du produit, mais directement du procédé ou processus de production, afin de détecter tout dysfonctionnement.

- **Le guidage et contrôle de robots**, de plus en plus présents dans les usines modernes. Les systèmes de visions sont souvent couplés à d'autres types de capteurs pour ce genre d'applications, afin de déterminer la position ou la trajectoire du robot par exemple.

Parmi ces catégories, c'est l'inspection visuelle automatisée qui regroupe la majorité des applications. De plus, certaines applications sont à cheval entre différentes catégories, notamment entre l'inspection visuelle, l'identification de produits et le contrôle de procédés. En effet, certaines applications en identification de produits ou en contrôle de procédés peuvent être basées sur des résultats d'inspection visuelle. Certains exemples d'applications pour chaque catégorie sont présentés en Figure 1.14.

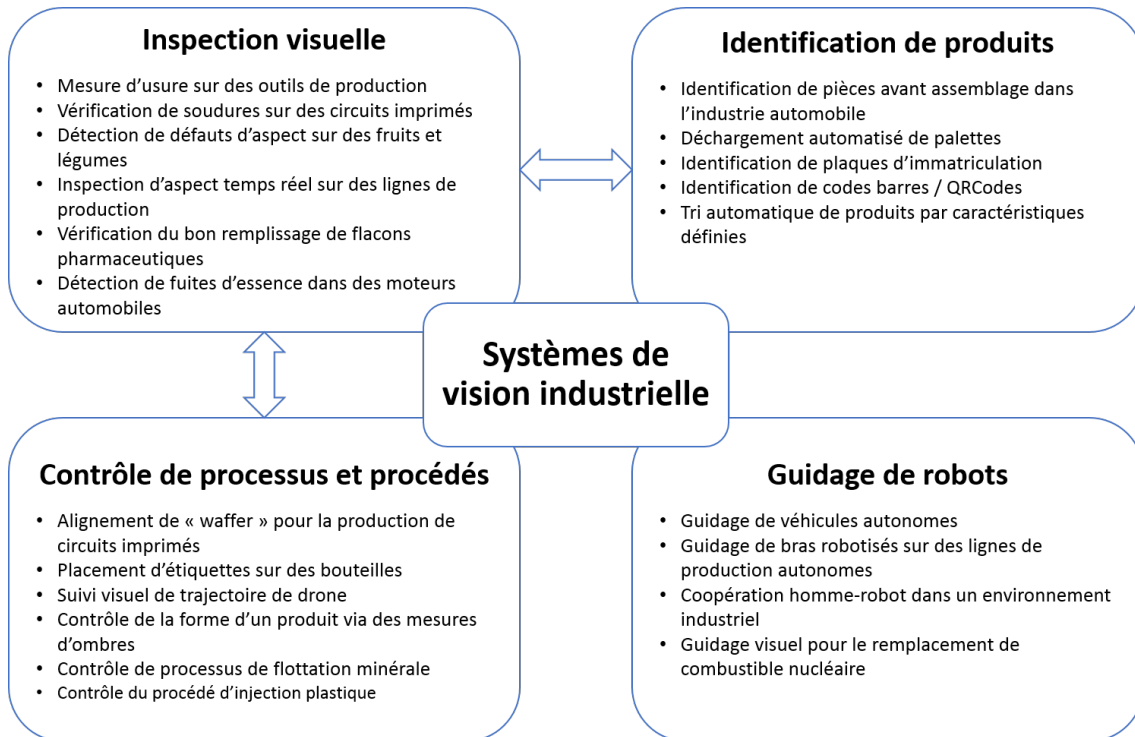


Figure 1.14 – Quelques exemples d'applications pour chaque catégorie de systèmes de vision industrielle selon la classification donnée par Golnabi

Classification axée sur l'inspection visuelle

Malamas propose également une classification guidée par l'objectif visé par le système de vision. Néanmoins, il se focalise pour sa part sur les applications dans l'inspection de la qualité des produits [Malamas2003]. Il divise ces applications en quatre classes basées sur le type de contrôle mené :

- **L'inspection de la qualité dimensionnelle** : vérifications des dimensions, formes, positionnement, orientation, alignement, circularité ou encore angles d'un produit. Il s'agit ici de s'intéresser aux caractéristiques géométriques en 2 ou 3 dimensions des produits.
- **L'inspection de la qualité de surface** : vérification de la rugosité, de la texture, de la teinte, ou encore de la présence de rayures, de fissures, de marques d'usure, de marques de collage, de plis, etc.
- **L'inspection de la qualité structurale** : vérification du bon assemblage de pièces entre elles au travers de la détection d'éléments manquants (trous, vis, clous, écrous, rivets, joints et autres éléments de fixations), mais également vérification de la présence d'objets étrangers non souhaités.

- **L'inspection de la qualité opérationnelle** : vérification du bon fonctionnement du produit par rapport aux standards et spécifications. Cette tâche est moins répandue pour les systèmes de vision et est de manière générale réalisée à l'aide d'autres capteurs spécifiques.

Malamas fait remarquer que si chaque catégorie porte sur des types de qualités différentes, elles ont toutes en commun de se réduire à une action de confirmation ou non que la pièce répond à des standards et spécifications données. Ces applications consistent donc dans la majorité de cas en une décision binaire *Conforme/Non Conforme*.

Classification axée sur les degrés de liberté (ddls) du système

Malamas propose également une autre classification des systèmes d'inspection industrielle orientée non plus vers l'objectif visé, mais sur ce qu'il appelle les "*degrés de liberté*" (ddls) du processus d'inspection. Ces ddls sont des paramètres du système dont l'état n'affecte pas le processus d'inspection, rendant donc le système indépendant à leurs variations. Ils permettent de caractériser le degré de flexibilité du système [Malamas2003]. Parmi les ddls les plus rencontrées en industrie se trouvent :

- La **position** des pièces inspectées (caractérise l'invariabilité en rotation ou translation).
- La **taille** (dimensions géométriques) des pièces inspectées.
- La **forme** des pièces inspectées.
- La **couleur** des pièces inspectées.
- La **texture visuelle** des pièces inspectées.
- Les **conditions d'illumination** des pièces inspectées.

Cette classification permet de caractériser la dépendance d'un système à l'application pour laquelle il a été conçu à l'origine. En effet, le moins de ddls un système possède, le plus il est dépendant de son application d'origine. Il est alors peu adaptable et peu évolutif.

C'est par exemple le cas du système proposé par Li, visant à inspecter la surface de barreaux métalliques [Li2012]. Le système d'inspection repose sur l'acquisition d'une image sous forme de matrice linéaire, via une source d'éclairage laser linéaire (Figure 1.15). Le dispositif d'acquisition ne possède donc pas beaucoup de ddls car l'objet doit être dans une position particulière (normale à faisceau), avoir une forme et taille particulière (cylindrique de 46 mm de diamètre) et l'illumination est très spécifique. Le système est donc performant dans l'application pour laquelle il a été conçu, mais sera peu adaptable à des variations d'applications (autres formes d'objets par exemple).

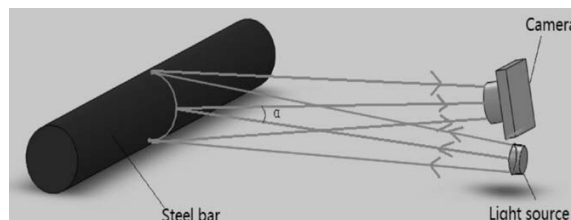


Figure 1.15 – Système d'acquisition pour l'inspection de barreaux métalliques [Li2012]

Au contraire, un système possédant un grand nombre de ddls, c'est à dire de variabilités ne gênant pas son fonctionnement, est moins dépendant de son application d'origine. Il possède alors un plus grand potentiel d'adaptation et d'évolution. De tels systèmes sont souvent plus compliqués à développer au départ car faisant souvent appel à des techniques plus sophistiquées d'acquisition d'images mais aussi d'analyse et la classification des données.

Le dispositif proposé par Leemans pour inspecter et classer des pommes en fonction de leur taille,

couleur, forme et présence de défaut est un exemple présentant un plus grand nombre de ddls [Leemans2004]. L'acquisition des images est réalisée dans un tunnel fermé, avec une illumination diffuse via des tubes fluorescents (Figure 1.16). Les fruits sont convoyés par un tapis roulant permettant également de faire rouler les pommes sur elles-mêmes. Le système d'acquisition est en effet peu impacté par la position des fruits par exemple, car ceux-là roulent afin d'être observés sous tous les angles de vue. La taille et la forme des fruits peuvent également être considérées comme des ddls car elles font parties des caractéristiques que le système doit inspecter. Elles sont donc amenées à varier dans une certaine mesure sans impacter le fonctionnement. Les conditions d'éclairages, bien que spécifiques dans un sens à l'application donnée, sont également impactantes dans une moindre mesure. Le système semble donc dans ce cas plus adaptable à des variations d'application (inspection d'autres types de fruits par exemple).

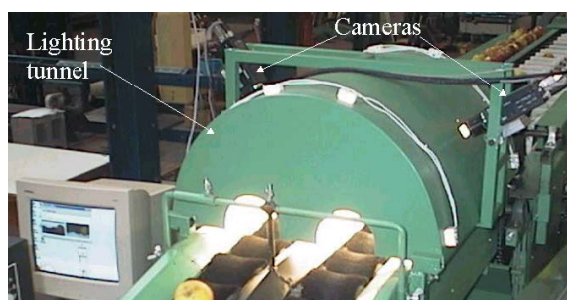


Figure 1.16 – Système d'acquisition pour le classement de pommes [Leemans2004]

La flexibilité d'un système, qui est définie par son niveau de dépendance à son application d'origine, reste néanmoins une notion relative. En effet, il apparaît clairement que celle-ci s'inscrit tout de même dans certaines limites fixées par l'application. Ainsi, une flexibilité dans tous les ddls reste difficile à atteindre, et une limite de variation reste présente pour chaque ddl. Par exemple, un système conçu pour inspecter des pièces d'une taille de l'ordre de quelques centimètres de côté, peut être très flexible en ce qui concerne la position et la forme des objets contrôlés ou les conditions d'éclairages, mais restera difficilement adaptable à des objets de tailles d'un ordre de grandeur supérieur. Ou encore, un système conçu pour inspecter des fruits et légumes de formes, tailles et couleurs différentes, restera difficilement adaptable à des produits présentant des caractéristiques surfaciques très éloignées comme du verre par exemple.

Ces considérations mettent en exergue qu'il n'existe pas de système de vision industrielle capable de gérer toutes les tâches dans tous les domaines d'applications. Elles mettent également en avant une problématique majeure dans la conception de systèmes de vision industrielle : le compromis à trouver entre flexibilité, complexité et coût. Un des grands challenges pour les développements futurs de l'inspection automatisée réside dans la conception de systèmes plus flexibles, restant efficaces dans un milieu industriel et présentant un coût raisonnable.

Après avoir exposé séparément ces différentes classifications, qui permettent d'obtenir un aperçu des applications et problématiques liées aux systèmes de vision industrielle, nous proposons de les synthétiser sous forme d'une seule et même représentation.

Synthèse des classifications présentées

La Figure 1.17 offre une synthèse des classifications présentées précédemment, afin de présenter un point de vue global des différents types de systèmes de vision industrielle. Elle intègre les trois classifications présentées précédemment. Ainsi, les systèmes sont tout d'abord classés selon leur domaine applicatif dans lesquelles on retrouve les 4 catégories : inspection de produits, identification de produits, contrôle de procédés/processus et guidage de robots. Les trois premières

catégories, qui regroupent la plupart des applications, présentent des liens privilégiés car certains systèmes sont à cheval entre elles. C'est par exemple le cas des systèmes de classification de produits basés sur une inspection visuelle.

La catégorie de l'inspection visuelle est ensuite décomposée en 4 autres classes en fonction de la nature du contrôle qui peut porter sur : la qualité dimensionnelle, la qualité structurelle, la qualité d'aspect ou encore la qualité opérationnelle d'un produit. Nous modifions ici légèrement la classification donnée par Malamas en introduisant le terme de qualité d'*aspect* au lieu de qualité de *surface*. En effet, la qualité de surface peut faire référence à la fois à des caractéristiques liées à l'aspect mais également à des caractéristiques géométriques. Une mesure de l'état de surface d'un produit fait par exemple écho à une mesure plus dimensionnelle qu'esthétique. Nous proposons donc de parler de qualité d'aspect et de préciser que la qualité dimensionnelle peut référer à différentes échelles (produit global ou zoom sur la surface).

Enfin, les degrés de libertés proposés par Malamas apparaissent comme une problématique englobant toutes les classes de systèmes. Ils permettent de qualifier la flexibilité des dispositifs. Aucune catégorie n'est clairement définie, mais plus un système en possède, plus il est considéré comme flexible et adaptable.

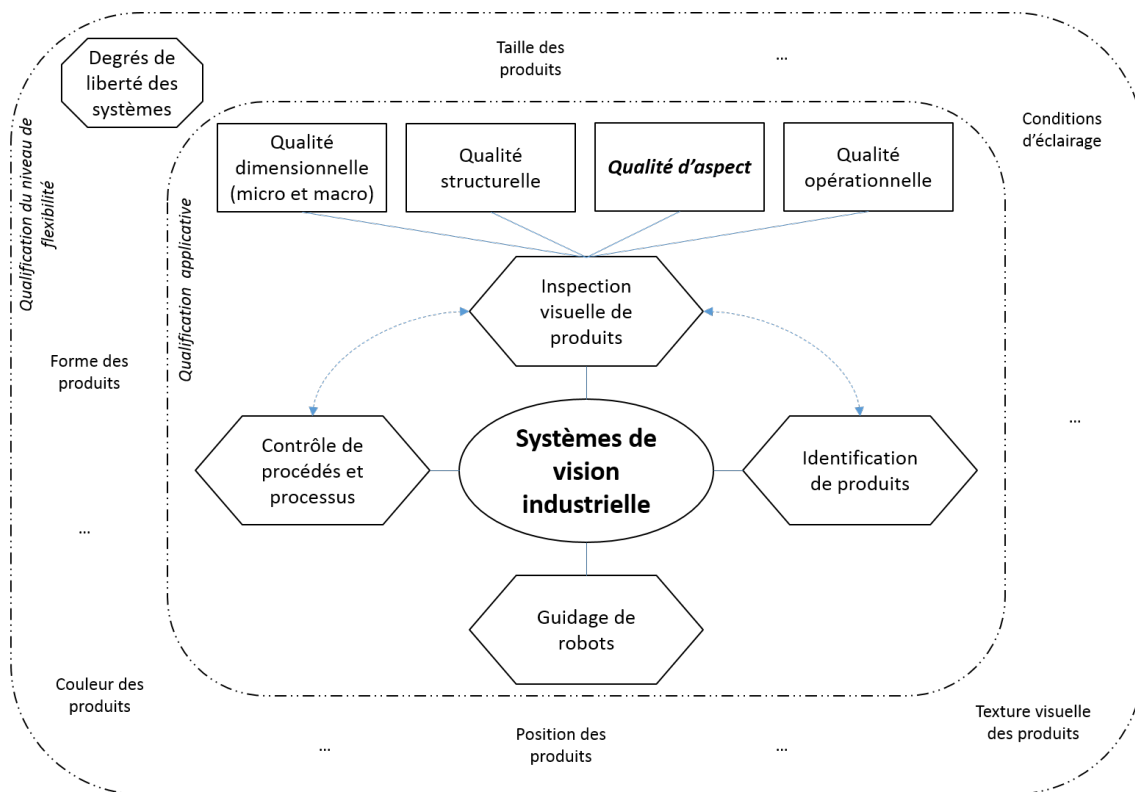


Figure 1.17 – Synthèse des classifications des systèmes de vision industrielle

Ces classifications permettent de prendre conscience de la diversité des applications des systèmes de vision industrielle. Ils permettent tout d'abord d'augmenter les performances des lignes de production, notamment en terme de contrôle qualité ou de contrôle des procédés, mais ils jouent également un rôle clef dans le développement de la flexibilité de ces dernières.

Le travail présenté dans cette thèse s'inscrit dans le développement des systèmes d'inspection de la qualité d'aspect des produits, catégorie que nous détaillerons un peu plus à la fin de cette partie, après la présentation de quelques points clefs pour le développement de systèmes de vision

industrielle.

1.3.1.2 Points clefs pour la conception d'un système de vision industrielle

L'étude des différentes applications au regard du modèle présenté dans la partie précédente en Figure 1.12 permet de dégager différents points clefs pour la conception d'un système de vision industrielle, présentés en Figure 1.18. Proposés à l'origine par Golnabi, nous proposons de les enrichir, notamment après l'étude des classifications précédentes.

Ces points clefs peuvent être décomposés en 6 catégories : le positionnement applicatif, les considérations liées à l'environnement de travail, les considérations systématiques, l'acquisition des images, le pré-traitement des images et la traduction et exploitation de l'information.

Le *positionnement applicatif* réfère tout d'abord à l'objectif visé par le système, première considération à prendre en compte pour la conception du système. Les différentes applications peuvent être catégorisées selon la classification détaillée précédemment.

Ensuite, il est également important de *considérer l'environnement de travail* dans lequel le système doit opérer, et notamment les contraintes qui y sont associées. Dans un environnement industriel, ces contraintes peuvent se traduire par exemple en terme de cadence à tenir ou encore dans le fait de devoir travailler avec un certain éclairage ambiant. Des contraintes peuvent également apparaître au niveau des produits à inspecter, en terme de taille, de forme ou de position lors du contrôle par exemple. Il est important de les prendre en considération pour concevoir un système de vision capable d'opérer de manière efficace dans son environnement.

Dans certains cas, ces contraintes peuvent se muer en avantages et être exploitées afin de faciliter l'inspection. Par exemple, si les zones d'intérêts à inspecter sont limitées, le dispositif peut être conçu pour se restreindre à ces zones et ainsi gagner en rapidité d'exécution.

Certaines *considérations systématiques* sont également importantes à prendre en compte. Elles portent par exemple sur le potentiel d'intégration du système dans son environnement et sur son potentiel d'industrialisation en vue d'une utilisation à grande échelle.

Le degré de robustesse souhaité pour le dispositif est également à considérer lors de sa conception, ainsi que son degré de flexibilité. Cette flexibilité est définie par le nombre de degrés de libertés laissés au système.

Ces considérations incluent également des réflexions à propos du type de connectivité à mettre en place. Le choix peut porter sur des connectiques génériques types USB par exemple, ou sur des connectiques plus spécifiques, en fonction de l'environnement de travail, du niveau de robustesse souhaité, ou encore du matériel utilisé.

Enfin, il est également essentiel de prévoir un système facile d'utilisation que les opérateurs pourront faire fonctionner sans difficultés.

L'*acquisition des images* joue un rôle clef dans les performances du système. Il est donc important de bien choisir le type d'éclairage utilisé pour commencer. Si beaucoup de développements optent pour des éclairages de type incandescent, halogène ou LED, il peut être également pertinent d'utiliser d'autres types de sources comme un laser ou d'autres types de rayonnements comme les rayons X, les ultraviolets (UV) ou les infrarouges (IR), en fonction de l'application visée.

Ensuite, le choix du capteur optique utilisé est également important, et peut être conditionné par la source d'éclairage choisie. Si beaucoup de développements optent maintenant pour des capteurs CCD ou CMOS, il existe d'autres capteurs plus simples comme les photodiodes ou les tubes vidicon par exemple. Si la source d'éclairage utilisée émet des rayonnements spéciaux (X, UV ou IR), le capteur utilisé devra être également choisi en conséquence.

Différentes configurations en terme de nombre et de position pour l'éclairage et l'acquisition op-

tique peuvent par la suite être considérées. Certains systèmes fonctionneront par exemple avec un point d'éclairage et un capteur, alors que d'autres nécessiteront plusieurs points d'éclairages et plusieurs capteurs.

D'autres considérations comme le filtrage des images directement sur le système d'acquisition d'images, ou le type de représentation utilisé (2D ou 3D) peuvent être également prises en compte. Le cinquième type de considération porte sur les éléments de pré-traitement des images. Il peut en effet s'avérer crucial de choisir les bonnes opérations afin d'améliorer l'information contenue dans les images. Il s'agit alors de bien identifier quels sont les caractéristiques des images acquises qui gagnent à être améliorées (contraste, balance des couleurs ou encore luminosité par exemple).

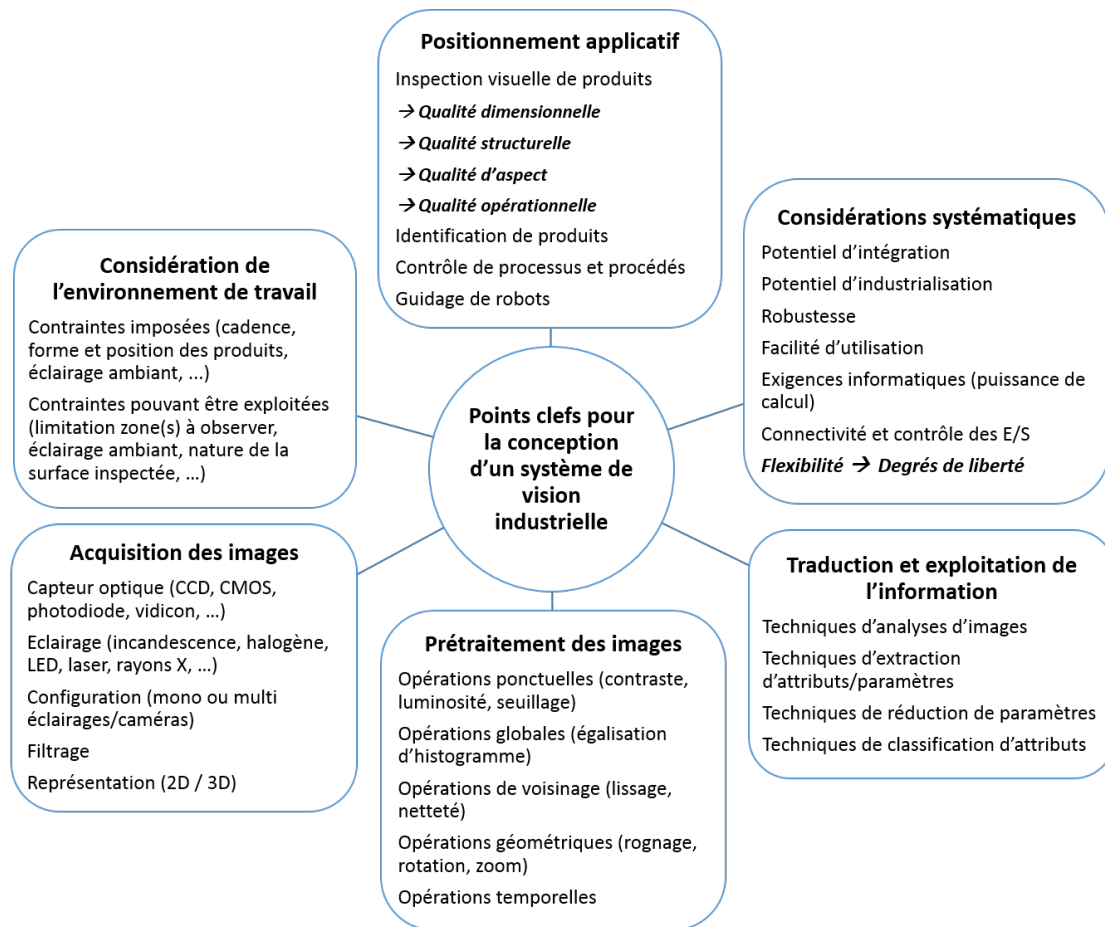


Figure 1.18 – Points clés pour le développement d'un système de vision industrielle, dérivés de [Golnabi2007]

Enfin, le dernier type de considérations listées concerne la traduction et l'exploitation de l'information contenue dans les images. Il existe là aussi de nombreuses techniques permettant d'analyser les images et d'en extraire des paramètres, de réduire les paramètres extraits afin de ne garder que l'information la plus pertinente, et d'interpréter ces informations en classifiant ces paramètres. Certaines techniques d'extractions de paramètres vont notamment être plus pertinentes que d'autres en fonction de l'information à laquelle le système doit être sensible.

Ces différentes considérations mettent en avant l'aspect multi-applications des systèmes de vision industrielle, ainsi que la difficulté à développer des systèmes très génériques. Pour la suite de cet état des lieux des systèmes de contrôle qualité automatiques par vision industrielle, nous proposons d'étudier le positionnement applicatif des systèmes proposés dans la

littérature tout d'abord par domaines d'activités, puis par rapport aux techniques employées, en nous focalisant sur l'inspection de la qualité d'aspect qui nous intéresse plus particulièrement.

1.3.2 Positionnement applicatif du contrôle automatisé de la qualité d'aspect

Pour faire suite à cette présentation générale des systèmes de vision industrielle, permettant de se rendre compte de leur variété ainsi que des problématiques liées à leur développement, nous souhaitons nous concentrer maintenant sur les systèmes dédiés au contrôle de la qualité d'aspect. Nous proposons dans un premier temps de donner un aperçu des secteurs d'activité d'application de ces systèmes. Nous axons ensuite cette analyse sur les principaux groupes de méthodes utilisés en terme de traduction d'images et de classification automatique. Ces deux angles d'étude permettent de mettre en avant l'état actuel des recherches dans ce domaine et certaines pistes de développement qui s'en dégagent.

1.3.2.1 Secteurs d'activités

L'automatisation de l'inspection visuelle d'aspect est une problématique répandue dans de nombreux domaines d'applications industrielles. Bien que les premiers développements remontent aux années 1970 avec des premiers états de l'art généraux dès les années 1980 par Chin [[Chin1982](#)] [[Chin1988](#)], nous souhaitons présenter ici un état des lieux plus récent (fin des années 1990 à aujourd'hui) de l'inspection automatisée de surfaces en industrie. Au vu du très grand nombre de publications présentées dans ce domaine, la bibliographie utilisée ne se veut pas exhaustive mais représentative de la variété des domaines d'applications rencontrés.

Nous présentons tout d'abord ici une répartition des dispositifs exposés dans la littérature en fonction de leurs secteurs d'application, afin d'identifier par la suite les domaines portant le plus d'intérêt pour de futurs développements.

Une répartition des publications en fonction des secteurs d'applications est illustrée en Figure [1.19](#), résumant de manière plus visuelle la répartition sous forme de liste du Tableau [1.3.2.1](#).

Le premier constat réalisé est qu'une grande majorité des communications référencées traite d'applications dans trois grandes industries : le textile, la métallurgie et l'agroalimentaire. Le reste d'entre elles se distribue entre divers industries comme celles des semi-conducteurs, des circuits imprimés (PCB), des écrans LCD ou encore de la plasturgie. On y retrouve aussi la présentation de techniques sans domaine d'application spécifique.

Partant de ce constat, il est possible de remarquer que les grands domaines d'applications des systèmes d'inspection automatisée d'aspect ont pour point commun de *grands volumes* de marchandises produites. En effet, les applications recensées sont souvent développées pour fonctionner sur des lignes de production présentant des temps de cycles courts, afin de répondre à des contraintes de productivité forte.

Ces contraintes vont souvent de pair avec une faible marge réalisée sur la vente des produits. Il est donc nécessaire de produire beaucoup en peu de temps pour garantir la rentabilité de l'activité. Dans ce cas, les entreprises présentent des moyens limités pour affecter des ressources humaines dédiées sur des tâches d'inspection de l'aspect des produits, et le recours à des solutions d'inspection automatisées se révèle donc avantageux.

De plus, même si la notion d'aspect des produits devient de plus en plus importante dans de nombreux domaines, elle ne constitue souvent pas la fonctionnalité principale des produits. Le contrôle visuel d'aspect fait alors parti d'un ensemble de différents contrôles et il est difficile de totalement dédier des contrôleurs à cette tâche.

Dans d'autres domaines, comme la joaillerie par exemple, l'aspect des produits présente une importance plus fondamentale, et il est ainsi plus commun de retrouver des contrôleurs dédiés à l'inspection d'aspect. De plus, les produits présentent généralement une plus grande marge à la vente, permettant de dégager plus de moyens pour rémunérer une main d'œuvre dédiée.

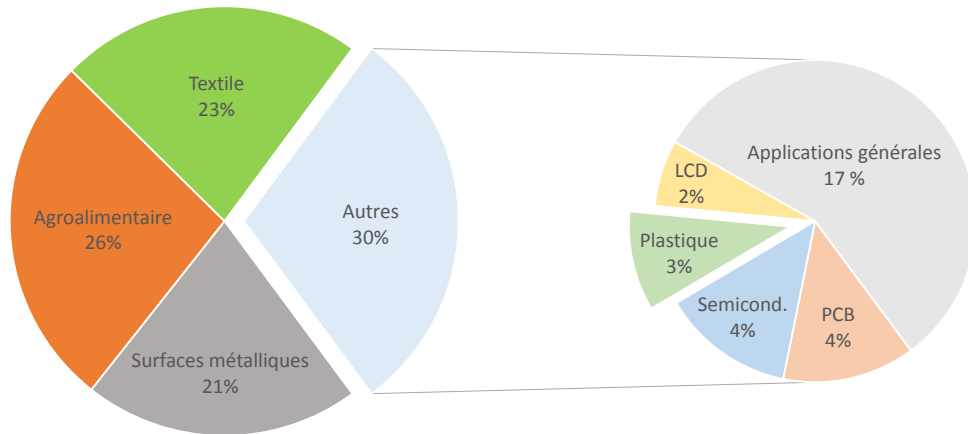


Figure 1.19 – Répartition par domaines d'activité des publications référencées un état de l'art récent composé de 125 publications référencées dans le tableau 1.3.2.1

Ainsi, le développement des problématiques de qualité d'aspect des produits dans des industries ayant peu de marge de manœuvre pour affecter des contrôleurs dédiés a ouvert d'importantes opportunités de développement de dispositifs automatisés dans ces domaines d'application.

Le second constat réalisé est que peu d'applications sont référencées dans l'industrie plastique, par rapport aux industries métallurgique, agroalimentaire et du textile. Pourtant, l'industrie plasturgique tient un rôle économique important, de poids comparable à ces 3 autres grandes industries, comme le montre les indicateurs présentés en Figure 1.20. C'est également un secteur en évolution croissante et à fort potentiel, notamment grâce au développement des techniques de fabrication additive, aussi connues sous la dénomination d'impression 3D.

Domaines	Surfaces métalliques	Agroalimentaire	Textile	PCB, Semiconducteurs, Plastique, LCD	Autres
Etat de l'art ou comparatif	[Neogi2014] [Pernkopf2003]	[Gunasekaran1996] [Zhang2014] [Brosnan2004] [Cubero2011]	[Kumar2008] [Ngan2011]	[Moganti1996] [Mar2011] [Huang2015]	[Chin1982] [Chin1988] [Newman1995] [Malamas2003] [Golnabi2007]
Articles présentant un cas d'application	[Blackledge2008] [Caleb2000] [Choi2007] [Clark Jr2001] [Hung2017] [Jeon2009] [Jeon2011] [Li2009a] [Li2012] [Luiz2010] [Mandriota2004] [Masci2012] [Pernkopf2003] [Pitard2017b] [Suvdaa2012] [Weiwei2008] [Wiltshi2000] [Yazdchi2009] [Zhao2010] [Zheng2002]	[Leemans2004] [Nicolai2006] [Baranowski2012] [Blasco2009] [Briones2006] [Diaz2004] [ElMasry2012] [Hu2014] [Khoje2012] [Li2013] [Majumdar2000] [Nagada2006] [Nicolai2006] [Peterson2007] [Quevedo2008] [Razmjoo2012] [Throop2005] [Unay2011] [Xiao-bo2010] [Xing2005] [Xing2007]	[Abouelela2005] [Baykut2000] [Bu2009] [Castellini1996] [Chan2000] [Cohen1991] [Conci1998] [Hajimowlana1998] [Kumar2003] [Kumar2000] [Kumar2002] [Kuo2003] [Liu2008] [Mak2008] [Mak2009] [Ngan2005] [Stojanovic1999] [Tolba1997] [Zhang1995] [Zhi2001]	[Chang2009] [Loh1999] [Lin2009] [Li2009b] [Shankar2006] [Kim2016] [Tien2013] [Zhou2014] [Lu2005] [Lu2008]	[Thomas1994] [Dennis1989] [Iivariinen2000] [Ko2000] [Latif-Amet2000] [Mäenpää2003] [Ng2006] [Rohrmus2005] [Tsa2000] [Tsai1999] [Xie2006]

Tableau 1.2 – Tableau répertoriant la bibliographie considérée, composée d'états de l'art, comparatifs et articles présentant des systèmes d'inspection automatisés dans les différents domaines d'applications

Secteurs d'activité* (extraits)	Nombre d'entreprises	Chiffre d'affaires HT Md €	Valeur ajoutée HT Md €	Taux de marge EBE/VA** (en %)
Réparation, installation et produits manufacturés divers	64 943	56,0	22,1	15,5
Industries agroalimentaires ⁽¹⁾	62 225	184,5	39,5	30,4
Bois, papier et imprimerie	34 617	38,9	10,8	16,0
Métallurgie et produits métalliques	22 562	83,4	25,2	16,0
Textile, habillement, cuir et chaussures	20 219	22,9	6,3	27,2
Plastique, caoutchouc et produits minéraux non métalliques	14 038	63,4	18,8	22,4
Machines et équipements	4 729	48,2	13,6	22,6
Industrie chimique	2 840	73,4	16,6	33,1
Produits informatiques, électroniques et optiques	2 832	30,2	10,6	13,9
Équipements électriques	2 617	28,0	8,0	17,0
Industrie automobile	2 041	101,5	16,1	12,4
Autres matériels de transport	1 026	52,0	15,9	26,4
Industrie pharmaceutique	346	38,3	10,7	20,8
Cokéfaction et raffinage ⁽²⁾	56	53,0	2,4	-25,6
Industrie manufacturière	235 092	868,9	216,5	21,3

* nomenclature NAF, niveau A38, sauf matériels de transport (A88).

** Ces taux de marge, par secteur d'activité, diffèrent de ceux des comptes nationaux, calculés par branche.

Champ : ensemble des entreprises marchandes y compris auto-entrepreneurs.

⁽¹⁾ hors produits à base de tabac.

Source : Insee, É sane 2014, sauf ⁽²⁾2013.

Figure 1.20 – Indicateurs structurels des secteurs industriels en France en 2015 [DGE2016]

Néanmoins, cette industrie présente quelques challenges pour l'automatisation de l'inspection. En effet, les pièces produites par une même usine présentent souvent une forte variabilité de taille, de forme ou encore de texture de surface et de couleur, notamment en injection plastique. En effet, une même ligne de production en injection sert généralement à créer une grande variété de pièces, les principales conditions pour passer d'un type de pièce à un autre étant de changer le moule d'injection ou/et le type de matière injectée.

Cette variété implique donc un besoin de flexibilité dans l'inspection de la qualité d'aspect des surfaces, pour s'adapter facilement aux différentes tailles, formes, textures ou encore couleurs des pièces injectées. Le développement de systèmes d'inspection automatisés flexibles est donc un enjeu important pour cette industrie.

Après s'être intéressés à la répartition des publications selon les domaines d'application, nous souhaitons maintenant analyser cette bibliographie du point de vue des moyens utilisés pour exploiter l'information acquise avec les images.

1.3.2.2 Moyens pour traiter l'information

L'objectif de cette étude est tout d'abord d'identifier les groupes de méthodes principalement utilisées dans la littérature, puis d'examiner si certaines méthodes sont utilisées de manière préférentielle.

Nous proposons de focaliser cette analyse sur deux étapes clefs des systèmes de vision industrielle : l'extraction de descripteurs et la classification.

Nous rappelons également que l'ensemble bibliographique utilisé ne se veut pas exhaustif, mais représentatif de la répartition des publications rencontrées.

Pour ce qui se rapporte à l'extraction de descripteurs², il est possible de diviser les méthodes utilisées en 4 catégories :

- **Approches géométriques ou de forme :**

Ces techniques portent sur une analyse morphologique des défauts permettant d'extraire des paramètres caractérisant leur forme (taille, centre, aire, moment d'inertie, etc.). On les retrouve dans les publications suivantes : [Blasco2009], [Caleb2000], [Clark Jr2001], [ElMasry2012], [Mak2009], [Razmjoooy2012], [Thomas1994], [Unay2011], [Xing2005], [Li2013], [Zheng2002].

- **Approches statistiques ou de distributions :**

Ces techniques exploitent la distribution des pixels d'une image et leurs interactions les uns avec les autres. Les auteurs suivants les utilisent pour traiter les images dans leurs systèmes de contrôle : [Caleb2000], [Chang2009], [Diaz2004], [Iivarinen2000], [Kumar2003], [Leemans2004], [Latif-Amet2000], [Li2009b], [Li2012], [Suvdaa2012], [Tien2013], [Unay2011], [Weiwei2008], [Zhao2010], [Zhou2014].

- **Approches par filtrages ou transformations :**

Ces techniques permettent d'exploiter l'information renvoyée par des filtres ou des transformations appliqués à l'image, comme celle de Fourier pour citer la plus connue d'entre elles. Les auteurs suivants mettent en avant l'utilisation de ce type d'approches : [Baranowski2012], [Chan2000], [Choi2007], [ElMasry2012], [Jeon2009], [Khoje2012], [Kumar2000], [Li2009a], [Lin2009], [Mak2008], [Mar2011], [Ng2006], [Peterson2007], [Tsa2000], [Wiltschi2000].

- **Autres approches par modèles :**

Nous regroupons dans cette catégorie des techniques variées, globalement moins référencées que les approches précédentes, utilisant divers types de modèles pour extraire des descripteurs. Parmi elles, il est possible de citer les approches par le modèle des fractals, par les champs aléatoires de Markov, par le calcul de la fonction d'autocorrélation, par la modélisation de la saillance de l'image, ou encore par réseau de convolution. Elles sont référencées par les auteurs suivants : [Baykut2000], [Blackledge2008], [Bu2009], [Conci1998], [Hung2017], [Ko2000], [Masci2012], [Nagada2006], [Pitard2017b], [Tolba1997], [Unay2011], [Xiao-bo2010].

Il est intéressant de noter que certaines publications apparaissent dans plusieurs catégories, comme celles de [Caleb2000], [ElMasry2012] et [Unay2011]. En effet, ces systèmes sont basés sur l'extraction de paramètres de différents types pour permettre une description plus flexible des anomalies. Par description plus flexible, nous entendons une description capable de détecter des anomalies aux caractéristiques variées. Néanmoins, ils restent minoritaires et la plupart des dispositifs proposés n'utilisent qu'un seul type de descripteurs, permettant de très bien décrire un type d'anomalie(s) ciblé. Cette observation rejoint les considérations exposées dans la partie précédente à propos de la difficulté à développer des systèmes de vision industrielle génériques.

Pour ce qui se rapporte à la classification des données³, il est possible de diviser les techniques en 2 grandes catégories :

- **Approches supervisées :**

Ces méthodes mettent en œuvre un apprentissage nécessitant un lot de données étiquetées par avance selon des classes définies par l'utilisateur. Elles sont utilisées par les auteurs suivants : [Baranowski2012], [Baykut2000], [Blackledge2008], [Blasco2009], [Caleb2000],

2. Les principes de l'extraction d'attribut et des techniques employées sont développés dans le Chapitre 2 de ce manuscrit.

3. Les principes de la classification de données et des techniques employées sont développés dans le Chapitre 4 de ce manuscrit.

[Caleb2000], [Conci1998], [Diaz2004], [ElMasry2012], [Hung2017], [Khoje2012], [Kumar2000], [Kumar2003], [Latif-Amet2000], [Li2009a], [Lin2009], [Mak2008], [Mar2011], [Masci2012], [Nagada2006], [Peterson2007], [Razmjoo2012], [Suvdaa2012], [Tsa2000], [Unay2011], [Weiwei2008], [Wilschi2000], [Xing2005], [Zhao2010], [Zheng2002].

- **Approches non-supervisées :**

Ces méthodes permettent de créer des classes automatiquement via un lot de données d'apprentissage non-étiquetées par l'utilisateur. Elles sont référencées en moins grand nombre que les approches supervisées et utilisées par les auteurs suivants : [Bu2009], [Caleb2000], [Chang2009], [Hu2014], [Iivarinen2000], [Ko2000], [Leemans2004], [Li2009b], [Tolba1997].

Du point de vue des méthodes de classification, les approches supervisées sont clairement plus utilisées. Néanmoins, certaines applications introduisent l'utilisation d'approches non-supervisées, quelques fois en complément d'une méthode supervisée, généralement afin de permettre au système de regrouper par lui même les données en différents groupes distinctifs. Les enjeux de l'utilisation de ce type d'approches sont explicités dans le chapitre 4.

Ce chapitre a permis de mettre en lumière certaines problématiques liées tout d'abord au contrôle visuel humain en industrie, puis aux systèmes automatiques existants pour répondre aux besoins d'automatisation dans ce domaine.

Les différents points mis en lumière dans ce chapitre nous permettent de définir un cahier des charges pour le développement d'un système d'inspection automatisé d'aspect inspiré du processus humain, dans un contexte industriel. Pour conclure cette partie, nous proposons donc de reprendre certains éléments essentiels identifiés précédemment pour constituer le cahier des charges qui a guidé ce travail de thèse.

1.3.3 Cahier des charges d'un système de contrôle reproduisant les caractéristiques spécifiques d'un contrôle humain

La première partie de ce chapitre a permis de mettre en avant que le contrôle visuel humain requiert une exploration de la pièce suivant différentes *scènes d'observation*, afin de maximiser la prise d'information permettant une détection optimale. Pour cela, les contrôleurs sont amenés à faire *miroiter* les pièces, c'est-à-dire faire varier l'angle entre la source de lumière, la surface et l'œil. Dans le cas du contrôle humain, la source de lumière et l'œil sont généralement fixes et le contrôleur fait bouger la surface de la pièce. Cela est plus compliqué à réaliser d'un point de vue vision industrielle car il s'avère compliqué de faire varier l'orientation des pièces sur une ligne de production. Néanmoins, il est possible d'atteindre le même résultat avec une pièce et un point d'observation fixes, en faisant varier les angles d'éclairage.

Ce type de dispositif d'acquisition multi-éclairage a déjà été utilisé dans la littérature, principalement dans le cadre d'applications de reconstruction d'aspect de surface et notamment dans les travaux de Le Goïc, Pitard et Desage, dans la continuité desquels s'inscrit notre travail de thèse [Le Goïc2012] [Pitard2016] [Desage2015]. Néanmoins, l'information multi-éclairage n'a encore jamais à notre connaissance été exploitée pour développer un système d'inspection automatisé sur des lignes de production.

Nous souhaitons donc proposer un système fondé sur l'exploitation de cette acquisition multi-éclairage, en montrant qu'elle contient une information pertinente pour la détection d'anomalies d'aspect.

L'intérêt de l'approche multi-éclairage est également de proposer un système flexible quant aux types d'anomalies à détecter. En effet, l'objectif est d'obtenir un système capable de détecter une grande variété d'anomalies.

Néanmoins, l'étude du processus de contrôle qualité d'aspect a également permis de mettre en évidence que la phase de décision finale résulte en une catégorisation des pièces inspectées comme

Conforme ou *Non conforme*. Nous proposons donc de baser notre approche sur une première phase de décision donnant l'information de conformité des pièces, suivie d'une seconde phase ayant pour objectif de catégoriser les pièces non conformes parmi les classes d'anomalies identifiées, avec un temps de calcul satisfaisant.

Une des difficultés majeures liée à l'exploitation d'une information multi-éclairage dans un contexte industriel réside dans la grande quantité de données à traiter. En effet, au lieu de traiter une ou quelques images comme dans les systèmes classiquement développés, l'acquisition multi-éclairage fournit quelques dizaines d'images. Afin de répondre aux contraintes de rendements des lignes de production industrielle, il est donc nécessaire de réduire l'information de manière efficace pour garantir un temps de calcul satisfaisant. La réduction d'information doit permettre à la fois de garder une information pertinente pour décrire la variété des anomalies considérées, mais également de réduire assez le temps de traitement pour satisfaire les contraintes de contrôle en milieu industriel.

Enfin, nous souhaitons également rendre le système flexible en terme de taille et de forme des objets à inspecter. En effet, l'étude du positionnement applicatif présenté précédemment a permis de mettre en évidence un enjeu de flexibilité des systèmes sur ces critères, notamment dans la perspective de développement dans l'industrie plasturgique. Pour cela, il est donc nécessaire de penser à une machine capable de s'adapter à des pièces de géométries variées.

Ce cahier des charges fait apparaître un certain nombre de défis de recherche que nous avons tenté de relever dans ce travail de thèse :

1. Mettre au point un système multi-éclairages (>30) compatible avec des temps de mesure sur ligne (quelques dizaines de secondes).
2. Identifier et mettre au point des algorithmes de réduction et de classification de l'information compatible avec le temps réel.
3. Être capable de détecter le plus possible de types de défauts sur des pièces plastiques variées.

1.4 Conclusion

Ce premier chapitre a eu pour objectif de présenter le contexte dans lequel s'inscrivent ces travaux de thèse. Nous avons exposé dans un premier temps les principes de fonctionnement et les problématiques liées au contrôle visuel humain, puis dans un second temps un état des lieux des systèmes de contrôle automatisés par vision industrielle. Enfin nous avons présenté le cahier des charges du dispositif de contrôle à inventer dans ce travail de thèse.

La première partie de ce chapitre a donc été dédiée à la compréhension du contrôle visuel humain, au travers notamment des résultats de travaux de recherches menés en amont de cette présente thèse au laboratoire SYMME.

Ils permettent de mettre en évidence la présence d'une variabilité intra et inter-contrôleurs lors de l'exercice du contrôle, qui s'avère problématique pour les entreprises industrielles. Afin de réduire cette variabilité, une modélisation ainsi qu'une formalisation du contrôle ont été proposées. Elles permettent notamment de réduire le vocabulaire employé pour désigner des anomalies d'aspect, mais également de séparer le contrôle en 3 étapes identifiées, l'exploration, l'évaluation et la décision, pour chacune desquelles des méthodologies standardisées sont associées. Ces dernières ont pour objectif de permettre aux contrôleurs d'inspecter les pièces de la même manière, réduisant ainsi les variabilités observées dans les résultats du contrôle.

Enfin, il a été question de mettre en parallèle ce modèle du contrôle humain avec un modèle d'inspection automatisé. En effet, le développement récent des technologies de vision numérique et de traitement de données, les solutions d'inspection visuelle automatisées se révèlent de plus en plus pertinentes pour aider à réduire la variabilité du contrôle. Il semble néanmoins important de prendre en considération l'expertise humaine dans le développement de ces systèmes, et nous avons donc présentés les principes d'une automatisation au regard du contrôle humain.

Dans la lignée de ces considérations, la seconde partie de ce chapitre s'est appliquée à présenter un état des lieux des solutions existantes pour le contrôle automatique par vision industrielle.

Les différentes classifications permettent de mettre en évidence la grande variété des applications visées par ce type de systèmes. En effet, il est possible de trouver des applications aussi bien pour des objectifs d'inspection visuelle, d'identification de produit, de contrôle de processus et procédés et de guidage de robots. Une grande partie des systèmes sont toutefois dédiés à l'inspection visuelle, qui peut elle-même se décomposer en l'inspection de la qualité dimensionnelle, de la qualité d'aspect, de la qualité structurelle et enfin de la qualité opérationnelle des produits.

Enfin, l'étude de ces différents systèmes de vision industrielle peut également être menée au regard des degrés de libertés qu'ils offrent, caractérisant leur degré de flexibilité. Ces considérations permettent de mettre en avant la difficulté à développer des dispositifs très génériques.

Nous avons ensuite concentré notre étude sur le positionnement applicatif des systèmes dédiés au contrôle d'aspect, par rapport aux secteurs d'activités d'application mais aussi des techniques mises en œuvre.

Du point de vue des secteurs d'application, il apparaît tout d'abord que le recours à des solutions d'inspection automatisées se révèle plus avantageuse dans des secteurs d'activités présentant une trop faible marge pour affecter des contrôleurs dédiés. Ainsi, la majorité des applications référencées se retrouvent dans les industries du textile, de la métallurgie et de l'agroalimentaire. Néanmoins, nous avons également mis en avant que peu d'applications sont répertoriées dans le domaine de la plasturgie, qui semble pourtant être un autre domaine d'application d'intérêt. Le développement de dispositifs pour ce secteur d'activité présente cependant un enjeu important de flexibilité pour s'adapter à la grande variété de pièces produites.

Du point de vue des moyens mis en œuvre pour traiter l'information, l'analyse bibliographique a permis de mettre en avant différentes catégories de méthodes utilisées pour décrire les images et classer l'information. Une analyse plus approfondie des techniques utilisées dans chacune des

catégories identifiées est proposée dans la suite de ce mémoire.

Enfin, à partir des considérations précédentes, nous avons exposé le cahier charges d'un système d'inspection inspiré du contrôle humain dans un contexte industriel, qui présente les caractéristiques principales suivantes :

- Exploiter une information multi-éclairage afin de mieux détecter les anomalies d'aspect.
- Garantir un temps de calcul correspondant aux attentes industrielles, impliquant réduction efficace de l'information.
- Classifier les pièces tout d'abord de manière binaire Conforme/Non Conforme, puis selon des classes d'anomalies dans la catégorie Non Conforme, en restant compatible avec une application temps réel.
- Garantir une flexibilité de fonctionnement du système d'inspection en terme de types de défauts, et de taille et forme des pièces à contrôler.

Les différents aspects de ce cahier des charges rejoignent le fil conducteur de la suite du travail présenté dans ce manuscrit. Nous aborderons en effet les moyens de réduire l'information de manière efficaces dans les chapitres 2 et 3, les méthodes permettant de classifier des données, notamment dans un contexte industriel, dans le chapitre 4, et enfin la pertinence de l'approche multi-éclairages et le développement d'un système flexible en terme de taille et de formes des pièces tout en tenant compte du contexte industriel dans le chapitre 5.

2

Réduction de l'information des images : Description

Sommaire

2.1	Introduction	42
2.2	Décrire une image : un état des lieux	43
2.2.1	Approches géométriques ou de forme	43
2.2.1.1	Représentation des images	43
2.2.1.2	Descripteurs géométriques	44
2.2.2	Approches statistiques ou de distributions	45
2.2.2.1	Approches de représentation	45
2.2.2.2	Descripteurs statistiques	50
2.2.3	Approches filtrage ou transformation	52
2.2.3.1	Domaine spatial	52
2.2.3.2	Domaine fréquentiel	52
2.2.3.3	Domaine spatio-fréquentiel	52
2.3	Décomposition modale discrète pour la description d'images	54
2.3.1	Principe de fonctionnement	55
2.3.2	Décomposition d'une géométrie	56
2.3.3	Application à l'analyse de l'aspect visuel des surfaces	58
2.3.4	Analyse multi-échelle par filtrage modal	58
2.3.5	Approximation de la surface de luminance	60
2.3.6	Description d'images via la DMD	62
2.4	Travaux applicatifs utilisant la description d'images	68
2.4.1	La DMD comme un indicateur de performance en Deep Learning	68
2.4.2	Description d'images pour prédire des caractéristiques haptiques	71
2.5	Conclusion	76

2.1 Introduction

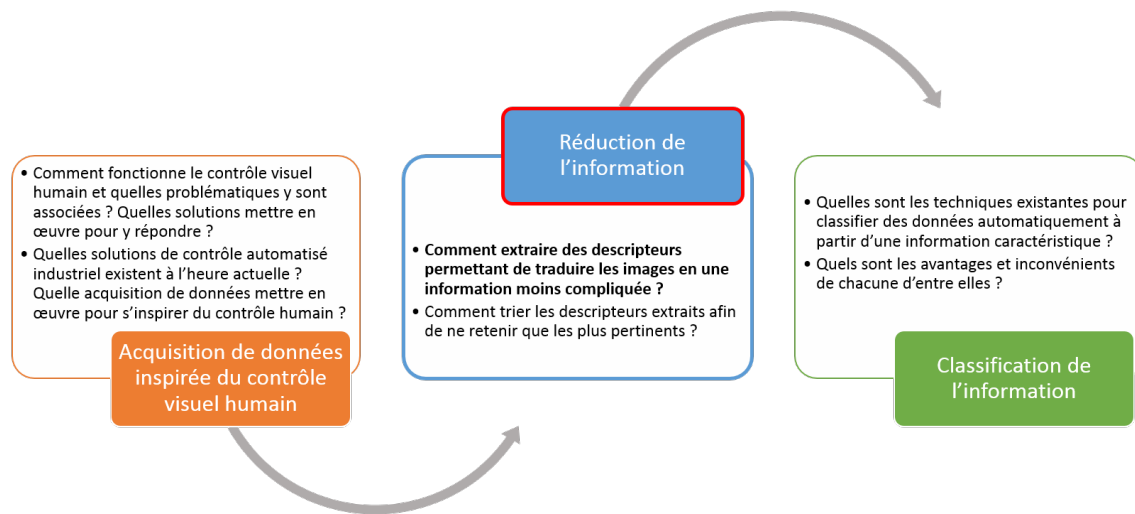


Figure 2.1 – Problématiques liées à la réduction de l'information par la description d'images

Les nombreuses images issues de l'acquisition multi-éclairages sont une information lourde à traiter, pouvant entraîner des temps de traitement non compatibles avec une application industrielle. Il est donc essentiel de réduire cette information afin de rendre son traitement plus efficace en terme de coup de calcul. Cette réduction doit néanmoins permettre de garder assez de pertinence dans l'information finale afin de pouvoir classer justement les objets par la suite.

Ainsi, il est nécessaire de mettre en œuvre une étape dite de **description des images**, ou **extraction de caractéristiques**. Elle a pour but d'extraire des *descripteurs*, ou *paramètres* des images, permettant de les caractériser avec une information moins lourde, pouvant prendre la forme d'une valeur seule ou un vecteur de valeurs.

Ce chapitre est tout d'abord dédié à présenter au lecteur les principes de fonctionnement de la description d'une image, en exposant notamment un tour d'horizon des techniques les plus utilisées dans la littérature.

Ensuite, nous introduisons l'utilisation d'une technique de paramétrisation de formes, la *Décomposition Modale Discrète*, développée lors de travaux précédents au laboratoire SYMME, pour décrire des images dans un but de classification. L'application de cette technique à l'extraction de descripteurs d'image dans un objectif d'apprentissage automatisé constitue un des apports originaux de ce travail de thèse.

Enfin, ce chapitre se termine avec la présentation de différents travaux applicatifs utilisant la description d'images et notamment la technique proposée précédemment.

De manière générale, ce chapitre cherche donc à répondre à la question : "Comment extraire des descripteurs permettant de traduire les image en une information moins compliquée ?".

2.2 Décrire une image : un état des lieux

La description d'image renvoie à un ensemble de techniques consistant en l'extraction de caractéristiques, appelées aussi descripteurs, ayant pour objectif de caractériser l'information contenue dans l'image.

En d'autres termes, cette étape permet de compresser les images en des représentations plus simples, tout en conservant l'information pertinente. Cette extraction de caractéristiques fait ainsi partie des processus de réduction de dimensionnalité.

La description d'une image utilise des méthodes permettant tout d'abord de représenter l'information contenue dans une image, et ainsi de mieux mettre en avant l'information qu'elle contient, puis des moyens de passer de ces représentations à des descripteurs pouvant être utilisés pour nourrir un algorithme d'apprentissage automatisé.

De manière générale, les différentes techniques peuvent être regroupées parmi 4 grandes familles [Tuceryan1993] [Porebski2009] :

- Les approches **géométriques** ou **de forme**,
- Les approches **statistiques** ou **de distribution**,
- Les approches **filtrage ou par transformation**,
- Les approches **par modèles**.

Cependant, il est à noter que la répartition des méthodes dans ces catégories n'est pas stricte et que certaines d'entre elles peuvent appartenir à plusieurs familles. Chaque type d'approche permet néanmoins d'analyser un type d'information spécifique contenu dans une image.

Dans le cadre de notre analyse, nous proposons de détailler les approches géométriques, statistiques, par transformation, qui sont les plus utilisées dans la littérature.

2.2.1 Approches géométriques ou de forme

Les approches géométriques, parfois aussi retrouvées sous le nom d'approches de forme, d'ensemble ou de région, permettent d'extraire une information *morphologique* de l'organisation de l'image. Elles permettent d'obtenir des descripteurs à partir de représentations fondées sur la segmentation de l'image en régions ou sur la détection de contours.

2.2.1.1 Représentation des images

La représentation des images pour cette famille de méthodes peut être réalisée suivant deux types d'approches [Desage2015] : les approches *forme* et les approches *contours*. Ces deux types d'approches permettent généralement d'obtenir une image binarisée caractérisant les régions¹ contenues dans l'image.

Approches contours

Ces approches permettent de caractériser les bords des régions par la détection des contours ou de points d'intérêt des différentes régions de l'image. Différentes techniques peuvent être utilisées

1. Nous définissons ici une *région* comme un ensemble de pixels ayant un ou plusieurs attributs en commun (valeurs identiques ou proches) ainsi qu'un lien spatial (connexité) qui font qu'on les distingue comme un ensemble dans l'image).

comme les filtres de Sobel, de Prewitt, de Canny ou encore l'opérateur Laplacien. Une illustration de la détection de contours de pièces est présentée en Figure 2.2.

Approches formes

Ces approches permettent de caractériser les formes contenues dans une image, à partir d'une version binaire de celle-ci. Cette image binaire peut représenter des contours, ou des régions complètes. Différentes techniques peuvent être utilisées pour caractériser les formes composant l'image, comme par exemple des méthodes d'approximations par enveloppe contenante. Elles consistent en la détermination d'une forme géométrique simple, comme un rectangle ou une ellipse par exemple, ajustée en terme d'amplitude et d'orientation à la forme d'une région de l'image. Une illustration de la détermination du rectangle englobant minimum pour deux formes définies par leurs contours est exposée en Figure 2.2. Les rectangles englobants sont représentés en blanc et les contours sur lesquels ils s'appuient en rouge.

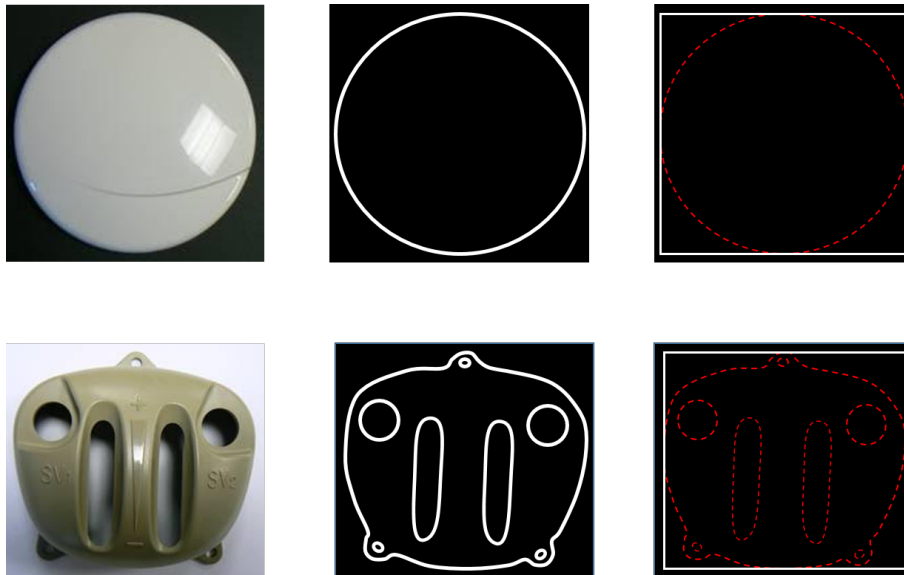


Figure 2.2 – Illustration d'une représentation du contour et du rectangle englobant minimum pour deux pièces de géométries différentes

A partir des représentations d'images par formes ou par contours présentées précédemment, des descripteurs dits géométriques peuvent être extraits.

2.2.1.2 Descripteurs géométriques

Les descripteurs extraits des représentations précédemment présentées sont dits *morphologiques* et donnent donc des informations concernant la géométrie des motifs composant l'image. Une liste non-exhaustive de ces descripteurs, reprise de [Desage2015], est :

- la **taille** correspond généralement à l'aire d'une région, c'est-à-dire le nombre de pixels qui la compose.
- le **périmètre** correspond au nombre de pixel composant les bords d'une région.
- la **direction principale** correspond communément à l'axe principal de l'ellipse englobante d'une région.

- la **densité** caractérise le nombre de *composantes connexes*² dans l'image binaire. Elle n'est intéressante que lors de la présence de plusieurs régions. Si il n'y a qu'une région, elle est égale à 1.
- le **nombre d'Euler** correspond à la soustraction du nombre d'objets connexes par le nombre de *trous*³.
- la **taille de l'enveloppe englobante** correspond à la taille, comme définie précédemment, non plus d'une région mais de l'enveloppe qui l'englobe. Il peut s'agir d'une enveloppe de forme elliptique ou encore rectangulaire.
- l'**excentricité** correspond au rapport des axes de l'ellipse englobant une région. Elle permet de caractériser si la région est plutôt morphologiquement proche d'une droite ou d'un disque.
- le **centre de gravité** correspond à la position du milieu de la région.
- la **distance moyenne** correspond de manière générale à la moyenne des distances des bords par rapport au centre de gravité.

Comme indiqué précédemment, cette liste n'est pas exhaustive. Le lecteur pourra trouver d'autres exemples de descripteurs géométriques dans un article de Pernkopf et O'Leary présentant un système d'inspection visuel de pièces métalliques cylindriques [Pernkopf2002].

2.2.2 Approches statistiques ou de distributions

Une image numérique est une matrice de pixels possédants chacun une valeur qui traduit l'intensité lumineuse reçue par le capteur visuel en chaque pixel. Certaines approches de traitement des images utilisent des mesures statistiques appliquées à la distribution ou aux interactions de ces valeurs pour caractériser une image. Ces méthodes font partie des premières à avoir été utilisées pour des applications de vision par ordinateur.

2.2.2.1 Approches de représentation

Si l'analyse statistique d'une image peut être faite directement sur les valeurs de l'image elle-même, il existe également différentes approches permettant de représenter l'information statistique ou de distribution d'une image. Nous souhaitons présenter ici les 4 plus utilisées dans la littérature, qui sont : les histogrammes, les matrices de cooccurrences, les motifs binaires locaux et les histogrammes des sommes et différences. Nous présentons l'application de ces outils à des images en niveaux de gris.

Histogrammes d'image :

Cet outil permet de représenter la distribution des intensités des pixels de l'image, sans information de localisation. Pour une image monochrome, le nombre de pixels correspondant à chaque niveau d'intensité est dénombré. Généralement, une image en niveaux de gris est constituée de 256 niveaux d'intensité, comme illustrée en Figures 2.4 et ???. Ce descripteur est très utilisé dans les applications d'inspection automatisée de part sa facilité d'utilisation.

2. Une composante ou objet connexe est un groupe de pixels de valeur 1 connectés entre-eux par leurs voisins directs

3. Un trou est un pixel ou un groupe de pixels à 0 entourés par un contour fermés de pixels à 1

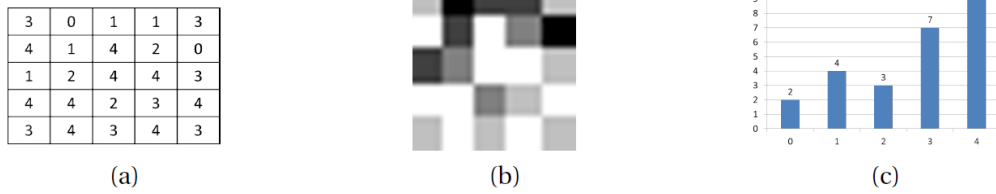


Figure 2.3 – Illustration d'une matrice carrée aléatoire de 5×5 pixels. Vue numérique (a), représentation en niveaux de gris (b) et visualisation de son histogramme (c). Images issues de [Desage2015].

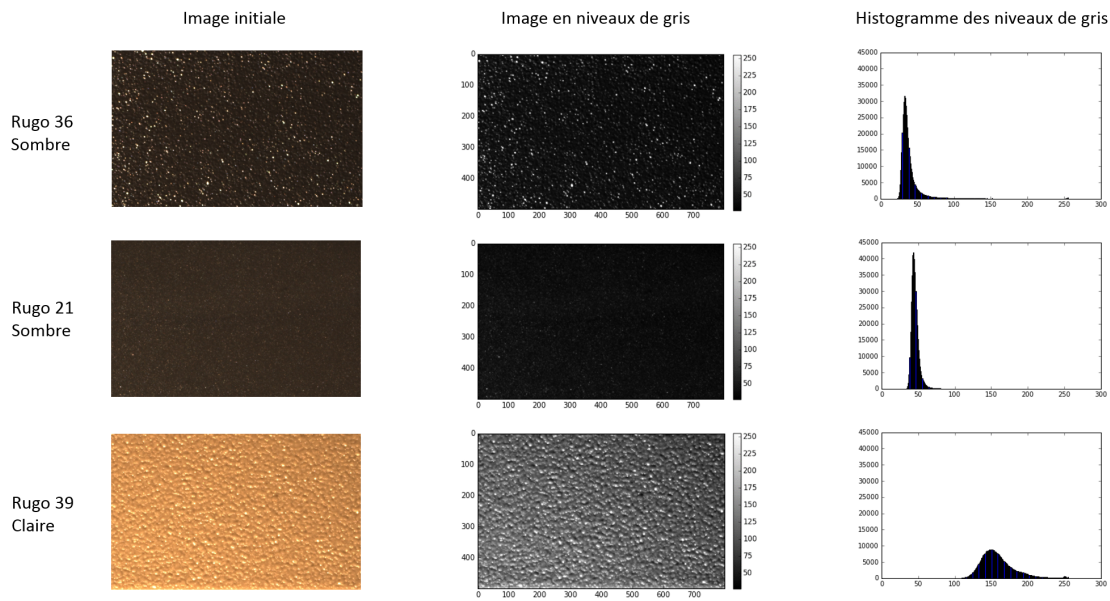


Figure 2.4 – Illustrations d'images de différentes surfaces rugueuses issues de plaquettes étalons charmillles, de leur transition en niveaux de gris, et enfin de l'histogramme associé. Les surfaces comportent différents niveaux de rugosité (21 = plus petit, 39 = plus grand) et différentes teintes

Les matrices de cooccurrences :

Cet outil, introduit originellement en 1973 par Haralick [Haralick1973], permet de caractériser la texture⁴ d'une image, communément en niveaux de gris. Chaque pixel est comparé avec ses voisins suivant un pas et une ou des directions données et le nombre de transitions d'une valeur de niveau de gris à une autre est compté. Les étapes du calcul pour des images en niveaux de gris sont les suivantes :

1. Choisir la valeur du pas et les angles des directions considérés ;
2. Pour chaque couple de niveaux de gris, compter le nombre de transitions présentes dans l'image ;
3. Reporter les comptes dans une matrice carrée $\text{Nombre de niveaux de gris} \times \text{Nombre de niveaux de gris}$ appelée *matrice de cooccurrences*.

4. La notion de texture est complexe et sa définition varie suivant le domaine d'application dans lequel elle est utilisée. Ici, nous la définissons comme un arrangement spatial de couleurs ou d'intensités, caractéristique d'une région de l'image. Elle dépend donc de l'échelle à laquelle on observe un objet.

L'illustration de la méthode de calcul est présentée en Figure 2.5 et 2.6.

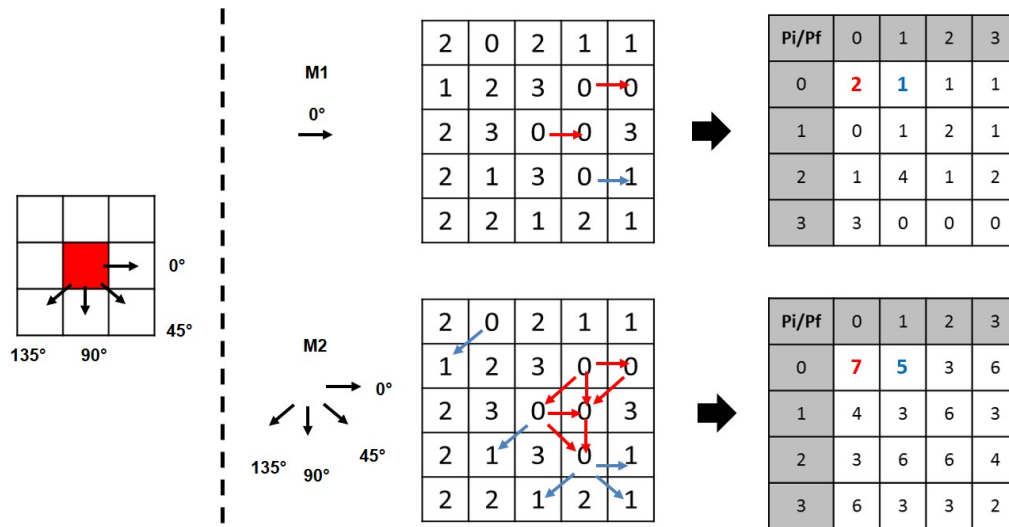


Figure 2.5 – Illustration du calcul de la matrice de cooccurrences d'une image 5×5 à 4 niveaux de gris, suivant 1 direction (M1) puis 4 directions (M2) et un pas de 1 dans les deux cas.

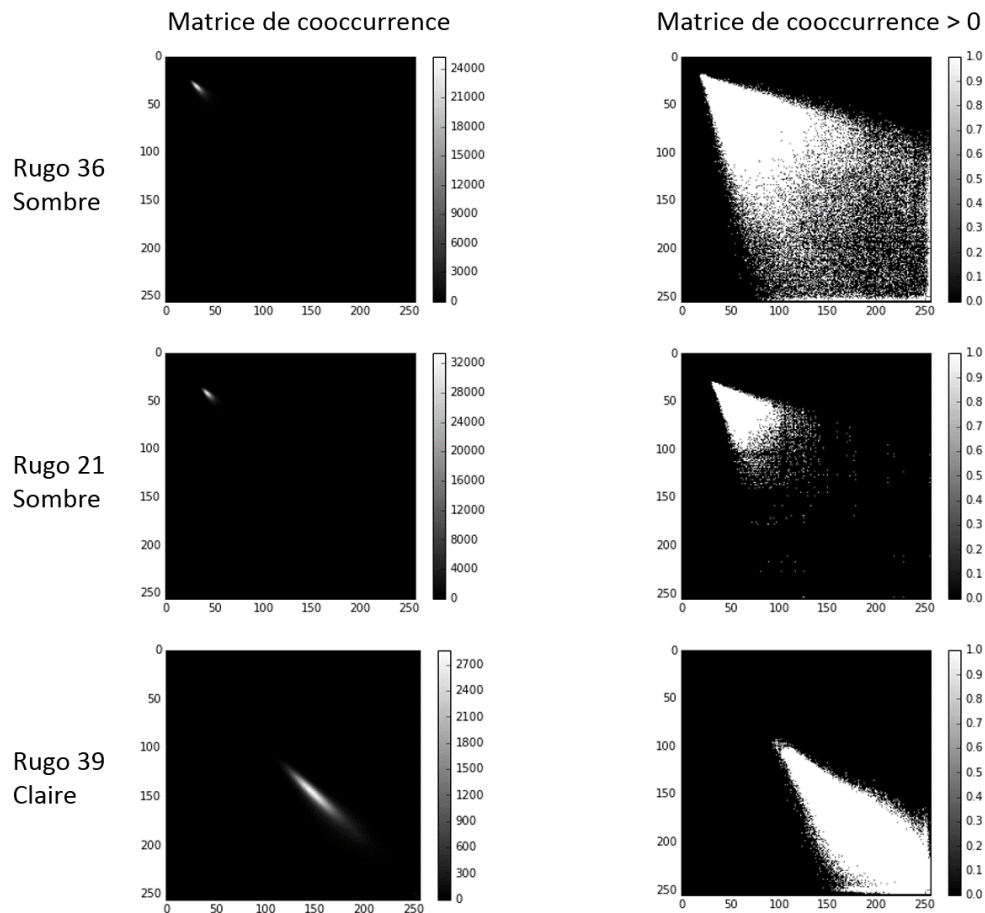


Figure 2.6 – Illustrations des matrices de cooccurrences des images précédemment présentées. Afin de mieux mettre en évidence les motifs des matrices, elles sont affichées avec un seuillage sur la colonne de droite.

Les motifs binaires locaux ou *Local Binary Patterns* (LBP)

Cet outil est un autre descripteur de la texture d'une image, utilisé dès 1996 par Ojala [Ojala1996]. Il fonctionne sur le principe de la comparaison de l'intensité d'un pixel avec celui de ses voisins. Chaque valeur de pixel est remplacée par une nouvelle valeur, calculée suivant les étapes suivantes :

1. Choisir l'échelle du voisinage utilisée (Figure 2.8) ;
2. Pour chaque pixel de l'image :
 - i) Calculer les différences entre les valeurs de chaque pixel voisin et le pixel considéré ;
 - ii) Attribuer une valeur 1 ou 0 à chaque pixel voisin en fonction du signe de la différence (0 si négatif, 1 si positif) ;
 - iii) Pondérer la matrice ainsi obtenue avec des facteurs 2^p , avec p le nombre de points de voisinage considérés ;
 - iv) Calcul de la nouvelle valeur du pixel considéré, aussi appelée descripteur du motif binaire, en sommant les valeurs de la matrice pondérée.
3. Reconstruction de l'image avec les nouvelles valeurs des pixels ainsi calculées.

L'illustration de cette méthode est présentée en Figure 2.7, et un exemple de la transformation d'une image au travers de celle-ci est exposée en Figure 2.9.

L'exploitation de l'histogramme (Figure 2.9) de cette nouvelle image peut par exemple permettre d'identifier différentes populations de motifs dans une image.

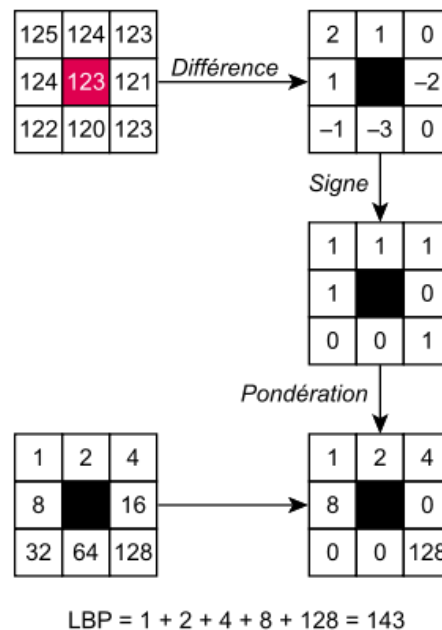


Figure 2.7 – Illustration du calcul de la nouvelle valeur d'un pixel pour un voisinage direct via la méthodes des LBP [CC : Par Xiawi — Travail personnel, CC BY-SA 3.0]

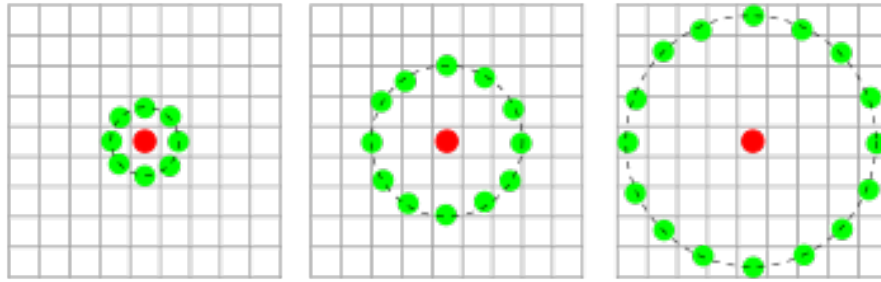


Figure 2.8 – Trois exemples de voisinages utilisés pour le calcul des LBP [CC : Par Xiawi — Travail personnel, CC BY-SA 3.0]

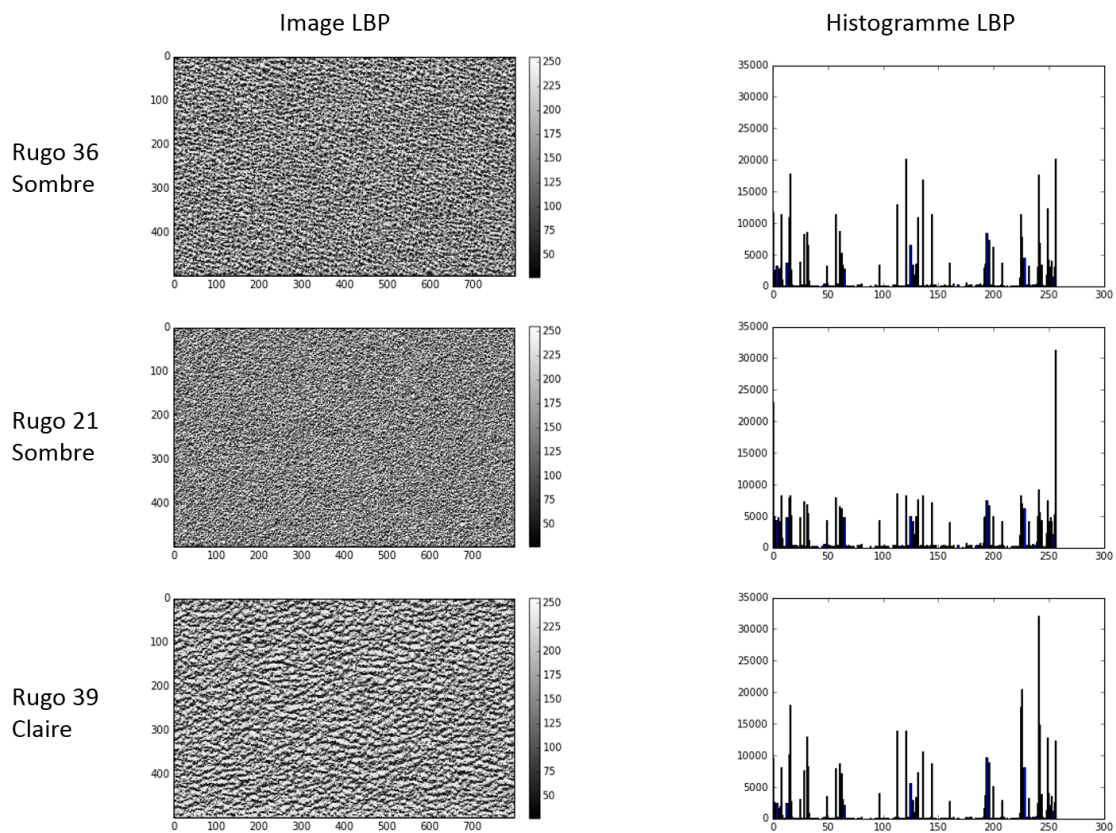


Figure 2.9 – Illustrations des images précédentes transformées par LBP. A gauche : les images après transformation. A droite : l'histogramme des images transformées.

Les histogrammes des sommes et des différences

Les histogrammes des sommes et des différences ont été initialement proposés par Unsero [Unser1986] comme une alternative aux matrices de cooccurrences pour l'analyse de la texture d'une image. Elles permettent également d'étudier la relation d'un pixel avec ces voisins et sont définies avec des directions et un pas de calcul, comme pour les matrices de cooccurrences. Leurs performances pour la classification de textures sont très proches de celles des matrices de cooccurrences et elles présentent l'avantage d'un temps de calcul plus court et moins gourmand en espace mémoire [Unser1986].

Le principe de calcul de ces histogrammes est présenté sur la Figure 2.10.

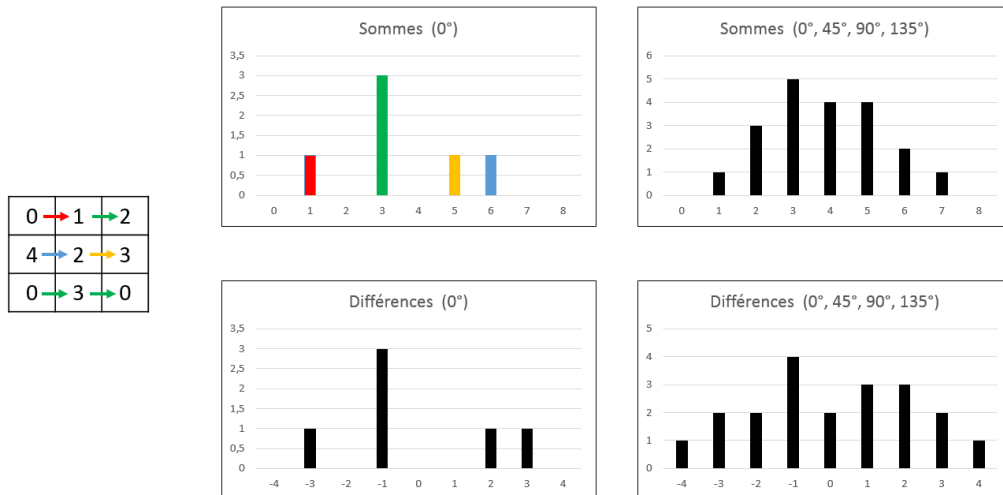


Figure 2.10 – Illustrations du calcul des histogrammes des sommes et des différences d'une image 3×3 pour un pas de 1 et un angle de 0° puis $0^\circ, 45^\circ, 90^\circ$ et 135° .

Bien que ces différentes méthodes de représentations permettent de réduire l'information contenue dans une image, les représentations obtenues ne sont communément pas utilisées directement pour alimenter un algorithme de classification. En effet, des descripteurs sont généralement extraits des histogrammes ou matrices afin de réduire encore le nombre d'information avec la classification.

2.2.2.2 Descripteurs statistiques

Dans le cas des approches de description d'images par distributions, des attributs dits statistiques sont utilisés. Ils peuvent être divisés en plusieurs catégories selon leur ordre. L'ordre d'un descripteur dépend du type d'interaction qui est considéré entre les pixels de l'image. Si on s'intéresse directement au pixel en lui-même, tel que dans l'histogramme classique, on parle d'ordre 1. Si on s'intéresse à des interactions entre couples de pixels, tel que dans la matrice de concurrence, les LBP et les histogrammes des sommes et des différences, on parle alors d'ordre 2.

Les **attributs d'ordre 1** peuvent être appliqués sur les histogrammes d'image (H) ou directement sur les valeurs des pixels de l'image (I). Parmi les plus utilisés, on retrouve :

- la **moyenne** (H, I).
- la **variance** ou l'**écart-type** (H, I), qui caractérisent la dispersion des valeurs des pixels autour de la moyenne.
- le **coefficient d'asymétrie** (H, I), ou *skewness* en anglais, qui mesure l'asymétrie de la distribution des pixels autour de la moyenne.
- le **coefficient d'acuité** ou d'aplatissement (H, I), ou *kurtosis* en anglais, qui mesure la concentration de la distribution des pixels autour de la moyenne. Il donne une indication de la concentration de la distribution à proximité (forme de dirac) ou à distance (d'une forme de gaussienne plus ou moins large, jusqu'à une forme de créneau) de la moyenne.
- la **médiane** (H), qui partitionne l'histogramme en deux populations égales.
- le **maximum** et le **minimum** (H, I). Dans le cas du maximum d'un histogramme, on parle aussi de **mode** ;

- certains **quantiles** (H) qui divisent l'histogramme en intervalles en un certain nombre de groupes contenant le même nombre de valeurs. La médiane en est un cas particulier pour la division en deux groupes égaux. Les quartiles (quatre groupes) et les centiles (cent groupes) sont d'autres quantiles communément utilisés.
- certains **intervalles inter-quantiles** (H) comme l'intervalle inter-quartiles par exemple, qui caractérise une dispersion autour de la médiane.
- l'**énergie** (I), calculée comme la somme des carrés des valeurs des pixels de l'image.
- le **contraste** (I), qui caractérise l'écart relatif entre la plus grande valeur de pixel et la plus petite dans l'image.
- l'**entropie** (I), qui caractérise la complexité de l'image, qui peut être traduite comme l'homogénéité des textures la composant.

Les matrices de cooccurrences et les LBP sont quant à eux exploitables grâce à des **attributs d'ordre 2** car leur calcul s'intéresse à des couples de pixels. Les plus utilisés ont été proposés par Haralick [[Haralick1973](#)] :

- | | |
|-----------------------------------|--|
| 1. le second moment angulaire | 8. l'entropie des sommes |
| 2. le contraste | 9. l'entropie générale |
| 3. la corrélation | 10. la variance des différences |
| 4. la variance (ou inertie) | 11. l'entropie des différences |
| 5. le moment différentiel inverse | 12. l'information sur la mesure de corrélation 1 |
| 6. la moyenne des sommes | 13. l'information sur la mesure de corrélation 2 |
| 7. la variance des sommes | 14. le coefficient de corrélation maximum |

Anys a également proposé l'utilisation d'autres attributs calculés sur les matrices de cooccurrences [[Anys1995](#)], qui ont notamment été repris par Verbrugge et Porebski [[Verbrugge2004](#)] [[Porebski2009](#)]. Les plus exploités sont :

- | | |
|---------------------------|-----------------------------------|
| 15. le moment diagonal | 20. la dissimilarité |
| 16. la moyenne | 21. la covariance |
| 17. le cluster shade | 22. la probabilité maximale |
| 18. le cluster prominence | 23. le troisième moment angulaire |
| 19. la différence inverse | 24. etc. |

D'autres descripteurs d'**ordre 2** reprennent certains attributs précédemment présentés pour les appliquer aux histogrammes des sommes et différences :

- | | |
|---------------------------------------|---|
| 17. le centre de gravité - somme | 25. le cluster prominence - différence |
| 18. la variance - somme | 26. l'entropie - différence |
| 19. le cluster shade - somme | 27. le second moment angulaire (ou énergie) |
| 20. le cluster prominence - somme | 28. le contraste |
| 21. l'entropie - somme | 29. la corrélation |
| 22. le centre de gravité - différence | 30. la variance |
| 23. la variance - différence | 31. le moment différentiel inverse : |
| 24. le cluster shade - différence | 32. l'entropie générale |

Le détail des formules de tous les descripteurs cités ci-dessus est disponible dans la thèse de Porebski [Porebski2009].

Après avoir présenté les familles de représentation et de descripteurs statistiques, nous proposons de nous intéresser maintenant à la famille des approches par filtrage ou transformation.

2.2.3 Approches filtrage ou transformation

Ces approches font références à un ensemble de techniques visant à transformer une image via des fonctions linéaires ou non. Elles permettent d'obtenir des coefficients, des mesures ou des statistiques à partir du filtrage de l'image au travers de ces transformations.

L'ensemble de ces approches peuvent être divisées en 3 catégories [Porebski2009] :

- Filtrage dans le domaine *spatial*,
- Filtrage dans le domaine *fréquentiel*,
- Filtrage dans le domaine *spatio-fréquentiel*.

Nous proposons de décrire brièvement les techniques les plus utilisées dans chaque catégorie.

2.2.3.1 Domaine spatial

Les approches de transformation dans le domaine spatial se résument principalement à l'utilisation de filtre pour la détection de contours [Tuceryan1993]. Ainsi cette catégorie se rapproche beaucoup de l'approche contour des descripteurs géométriques. On y retrouve donc généralement les filtres de Sobel, de Prewitt, de Canny ou encore les opérateurs de Roberts et Laplacien. Les masques associés à ces différents filtres et opérateurs sont présentés en Figure 2.11.

$$\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \quad \begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix} \quad \begin{bmatrix} -1 & 0 & -1 \\ -1 & 0 & -1 \\ -1 & 0 & -1 \end{bmatrix} \quad \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}$$

Figure 2.11 – Exemple de masques associés à différents types de filtres et opérateurs. De gauche à droite : Roberts, Laplacien, Prewitt, Sobel

2.2.3.2 Domaine fréquentiel

L'analyse dans le domaine fréquentiel tend à décrire une image en terme de *structures périodiques*, en la décomposant dans une base de fonctions périodiques simples. Les transformées De Fourier discrètes ou en Cosinus discrète sont principalement utilisées pour réaliser ces opérations. Le résultat de la décomposition permet d'obtenir des coefficients fournissant des informations fréquentielles sur le contenu de l'image. Elles sont notamment utilisées dans le cas de l'analyse d'images présentant des textures aux motifs répétitifs.

2.2.3.3 Domaine spatio-fréquentiel

Si les transformées Cosinus et de Fourier permettent une analyse globale du contenu fréquentiel de l'image, elles ne permettent pas de le localiser dans le domaine spatial. Néanmoins, il existe

des méthodes permettant de coupler les deux domaines, permettant ainsi d'analyser une image dans le domaine dit *spatio-temporel*. Parmi les transformations les plus utilisées, on retrouve :

- la transformation de Gabor,
- la transformation en ondelettes.

Ce type d'attributs sont également intéressant car ils se rapprochent du processus de la vision humaine qui réside entre partie sur une décomposition spatio-fréquentielle systématique des images [Campbell1968].

Nous proposons de les présenter rapidement dans les paragraphes suivants.

Transformation de Gabor :

Comme précisé précédemment, la transformation de Fourier pose une difficulté dans le fait qu'elle agit globalement sur l'image et ne tient pas compte de la localisation spatiale.

Une solution courante pour résoudre ce problème réside dans l'utilisation de la transformation de Fourier à fenêtre glissante. Le principe de cette transformation consiste en l'application de la transformée de Fourier dans une fenêtre d'analyse qui se déplace dans l'image. L'aspect spatial apparait donc grâce à la définition de cet espace d'analyse car il devient possible de localiser l'information fréquentielle. Le choix de la taille et du pas de déplacement de la fenêtre dépend généralement des caractéristiques spatiales des textures contenues pas l'image.

Il existe plusieurs types de fenêtres d'observation, comme par exemple : la fenêtre rectangulaire (fonction porte), la fenêtre triangulaire, la fenêtre de Welch, la fenêtre de Hamming, la fenêtre de Hanning, ou encore la fenêtre de Gauss. Leur visualisation graphique est présentée en Figure 2.12. Lorsque la fenêtre de Gauss est utilisée, on parle alors de *transformation de Gabor*, ou *filtrage de Gabor*.

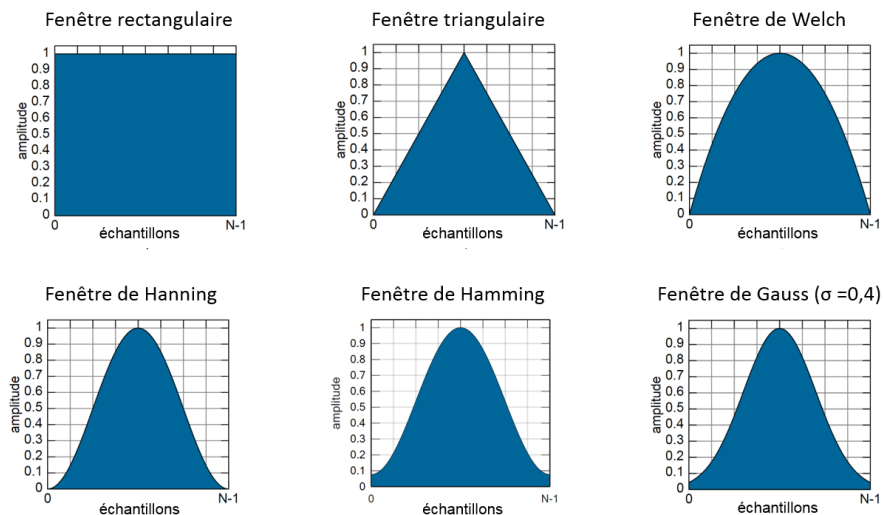


Figure 2.12 – Illustrations de quelques fenêtres utilisées pour la transformation de Fourier à fenêtre glissante.

Le filtrage d'une image au travers de cette stratégie permet alors de récupérer des coefficients pour chaque fenêtre. Pour un filtre donné, le descripteur caractérisant toute l'image est ensuite obtenu en calculant l'énergie, l'entropie ou la variance des coefficients obtenus au travers de chaque fenêtre [Porebski2009].

Il est cependant à noter que le réglage des paramètres du filtre n'est pas aisé, et que le temps de calcul peut être assez élevé, surtout pour des tailles de fenêtres élevées.

Transformation en ondelettes :

Une des limites de la transformation de Gabor est qu'elle s'effectue avec une longueur de fenêtre fixe. En effet, une image est souvent composée de textures pouvant être caractérisées suivant différentes échelles.

Afin de remédier à ce problème, la transformée en ondelettes introduit l'utilisation de plusieurs fenêtres de tailles différentes, afin d'effectuer une analyse multi-échelles de l'image. Elle permet ainsi de considérer à la fois des informations spatio-fréquentielles globales et locales composant l'image.

Cette transformation a également été rendue célèbre par son application pour la compression d'images, notamment avec le format JPEG 2000 [[Antonini1992](#)].

Ces trois catégories d'approches regroupent les méthodes les plus utilisées pour la description d'images appliquée notamment à la vision par ordinateur. Comme énoncé en début de cette section, il existe une quatrième catégorie que nous ne développons pas dans ce mémoire. Elle regroupe des techniques s'appuyant sur l'utilisation de modèles comme les champs aléatoires de Markov ou les fractales pour les plus utilisées. Ces méthodes peuvent être utilisées pour décrire des textures, mais également pour en synthétiser [[Tuceryan1993](#)].

Il est également important de faire remarquer ici que nous avons précédemment exposé des démarches de descriptions appliquées à des images en niveaux de gris, car elles sont plus simples et plus rapide à traiter. Néanmoins, il existe également des pendants de ces techniques pour l'application à des images en couleur. Il est ainsi possible de citer les matrices de cooccurrences chromatiques ou encore les motifs locaux binaires couleur [[Porebski2009](#)]. Elles consistent généralement à appliquer les même processus sur les différentes images-composantes et de regrouper les informations ainsi obtenues en des descripteurs globaux.

Après avoir présenté les principales méthodes habituellement utilisées dans la littérature pour décrire une image dans le domaine de la vision par ordinateur, nous souhaitons introduire l'utilisation d'une nouvelle méthode basée sur la décomposition de l'image via une technique de paramétrage descriptif, la Décomposition Modale Discrète. L'application de cette stratégie à la description d'images pour un objectif de classification constitue un des apports de ce travail de thèse.

2.3 Décomposition modale discrète pour la description d'images

Comme rappelé dans le chapitre précédent, ces travaux de thèse s'inscrivent dans la lignée de différents travaux au laboratoire SYMME, portant notamment sur le contrôle qualité d'aspect mais également sur le tolérancement géométrique de produits.

Parmi ces travaux, certains ont porté sur le développement d'une nouvelle méthode de paramétrage descriptif de surface, nommée **Décomposition Modale Discrète (DMD)**. Elle a notamment été appliquée à la modélisation et la reconstruction de l'aspect de surface de produits à haute valeur ajoutée [[Le Goïc2012](#)] [[Pitard2016](#)] [[Desage2015](#)].

Dans le cadre de mes travaux, je souhaite proposer une extension de ce travail, en utilisant cette méthode pour décrire l'information contenue dans une image.

Les origines et les principes de fonctionnement de cette technique sont présentées dans un premier temps, puis nous illustrerons son application à la description d'images.

2.3.1 Principe de fonctionnement

La DMD fait partie de la famille des techniques de *paramétrages descriptifs* de surfaces. Celles-ci ont pour objectif commun de décomposer une surface dans un espace de descripteurs, afin de caractériser la participation de chacun d'eux dans la forme géométrique complète. L'espace des descripteurs varie d'une technique à l'autre.

Parmi les méthodes les plus utilisées dans la littérature, on retrouve :

- La *Décomposition en Cosinus* basée sur la transformée discrète en cosinus (DCT).
- La *Décomposition en séries de Fourier* basée sur la transformée de Fourier discrète (DFT).
- La *Décomposition en harmoniques sphériques* basée sur un espace d'harmoniques sphérique et particulièrement utilisée pour la description de formes complexes en 3D.

La DMD est quant à elle basée sur un ensemble de descripteurs issus de la mécanique vibratoire, appelés *modes* ou *déformées modales*. Cette méthode est notamment issue des travaux de [Samper2007] et [Favreliere2009] dans le domaine du tolérancement géométrique. Elle a été utilisée depuis dans différents domaines d'application :

- Dans le domaine du **tolérancement géométrique**, pour caractériser et quantifier des variations et des écarts de formes [Samper2009] [Favreliere2007] ;
- Dans le domaine des **états de surfaces**, pour mener des analyses multi-échelles permettant de caractériser des défauts de forme, d'ondulation et de rugosité, pour des applications de contrôle de pièces à haute valeur ajoutée [Goic2011] et de prothèses chirurgicales [Grand-jean2012] ;
- Dans le domaine de l'**étude de champs thermiques**, pour évaluer la position de sources thermiques [Pottier2014] ;
- Dans le domaine de la **modélisation d'aspect de surface**, pour approximer la fonction de réflectance d'une surface afin de reconstruire un rendu réaliste permettant de mettre en évidence des défauts d'aspect [Pitard2015] [Pitard2017a].

L'intérêt de cette méthode pour l'analyse de l'aspect de surfaces réside dans le fait qu'elle puisse s'appliquer à tout type d'élément géométrique de référence (du plan jusqu'à une pièce spécifique complexe), et qu'elle permet de caractériser à la fois des géométries locales et globales ainsi que des éléments périodiques et non-périodiques. Cette propriété est très intéressante en analyse d'aspect de surfaces car celle-ci est généralement à la fois composée d'éléments périodiques ou pseudo-périodiques (éléments de forme globale ou de rugosité), et d'éléments non-périodiques (anomalies ponctuelles ou locales par exemple).

Partant des travaux et applications cités ci-dessus, nous proposons dans ces travaux d'étendre l'utilisation de la DMD dans le domaine de l'apprentissage automatisé, comme méthode de description d'une image.

2.3.2 Décomposition d'une géométrie

Comme les transformations de Fourier discrète ou encore en Cosinus, la DMD décompose un signal dans un espace de descripteurs définis *a priori*. La définition *a priori* induit que les descripteurs sont décorrélés du signal décomposé, c'est-à-dire qu'ils ont été générés en amont sans connaissance *a priori* du signal à décomposer.

Dans le cas de la DMD, le signal décomposé correspond à une surface discrète, pouvant provenir d'une mesure directe ou d'une simulation.

Base des descripteurs ou base modale

Les descripteurs, aussi appelés modes, peuvent dériver de tout type de géométrie de référence, que ce soit une forme élémentaire "simple", comme un plan, un cylindre ou encore une sphère, ou d'autres formes moins régulières. Leur détermination est obtenue à partir d'une analyse modale de l'élément de référence choisi. Les descripteurs sont alors les modes propres issus de l'analyse. La modélisation mécanique de l'élément de référence est caractérisée par une matrice de raideur K et une matrice de masse M . La détermination des modes propres du système revient alors à résoudre un problème de mécanique vibratoire défini par l'équation :

$$K \cdot q + M \cdot \ddot{q} = 0 \quad (2.1)$$

avec q le vecteur déplacement.

Il est alors possible de montrer que la détermination des modes propres Q_i se fait en résolvant le système linéaire décrit par l'équation :

$$(K - \omega_i \cdot M) \cdot Q_i = 0 \quad (2.2)$$

avec ω_i la pulsation liée au mode propre Q_i .

L'ensemble des modes Q_i ainsi obtenus est rassemblé dans une base de descripteurs nommée *base modale*. Ils sont organisés par ordre croissant des ω_i pour obtenir les modes les plus simples en premiers.

Pour permettre au lecteur de se représenter visuellement le résultat de ces opérations, les premiers éléments de la base modale calculée à partir d'un élément géométrique plan sont présentés en Figure 2.13.

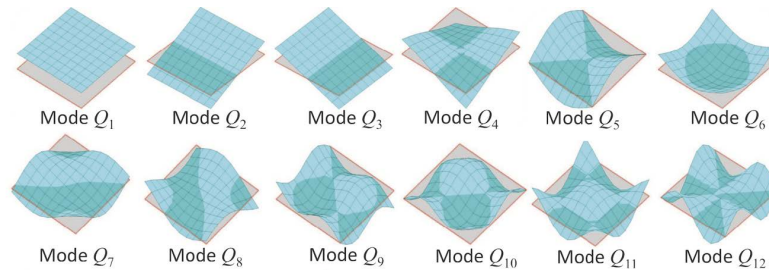


Figure 2.13 – Premiers modes de la base modale calculée pour un élément géométrique de référence plan [Le Goïc2012]

A partir de cette base modale, il est possible de décomposer des surfaces dans cet espace, afin d'analyser la participation de chacun des modes aux géométries de celle-ci. Cette opération de décomposition est décrite ci-après.

Projection d'une géométrie dans la base modale

L'opération de décomposition de la surface mesurée consiste en sa projection dans la base modale calculée au préalable. Si on nomme ^{mes}V le vecteur correspondant à la mesure de la surface, la projection se résume en l'expression de ^{mes}V sous la forme d'une combinaison linéaire des vecteurs Q_i de la base, décrite par l'équation :

$$^{mes}V = \sum_{i=1}^N \lambda_i \times Q_i + \varepsilon(N) \quad (2.3)$$

avec N le nombre de modes de décomposition choisis et λ_i les coordonnées du vecteur ^{mes}V dans la base des modes Q_i , aussi appelés *coordonnées modales* ou *coefficients modaux*. $\varepsilon(N)$ est le résidu de la projection, explicité un peu plus loin dans cette section par l'équation 2.6.

L'ensemble des coordonnées modales peuvent être regroupées sous forme d'un vecteur λ , obtenu via l'équation :

$$\lambda = Q^* \cdot ^{mes}V = (Q^T \cdot Q)^{-1} \cdot Q^T \cdot ^{mes}V \quad (2.4)$$

Afin d'obtenir les coordonnées modales en valeur métrique, homogène avec la surface mesurée, les vecteurs Q_i sont normés avec la norme infinie $\|Q_i\|_\infty = 1$. Ainsi, dans le cas des applications à la caractérisation d'écarts de formes par exemple, il est possible de quantifier les écarts par rapport à chaque mode.

Les vecteurs de coordonnées modales peuvent être représentés sous forme graphique pour analyser rapidement quel mode participe ou non à la composition d'une surface. On parle alors de *spectre modal* dont un exemple est présenté en Figure 2.14.

Les premiers modes caractérisent la participation d'éléments géométriques correspondant à des formes

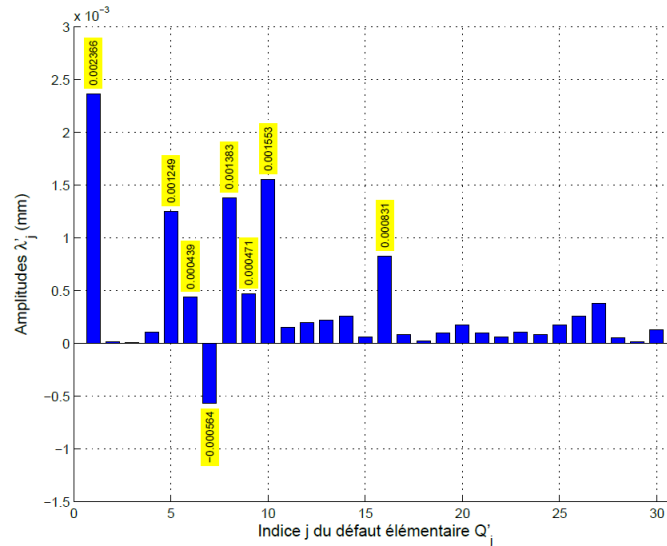


Figure 2.14 – Exemple de spectre modal

Le spectre modal permet donc d'observer les contributions correspondantes à chacune des formes élémentaires de la base modale considérée. Il permet donc de *paramétrer* une surface et constitue donc une signature propre de cette dernière.

Nous proposons maintenant de présenter succinctement la pertinence de cette approche pour l'analyse d'aspect visuel des surfaces.

2.3.3 Application à l'analyse de l'aspect visuel des surfaces

Dans le cadre du développement d'un système d'aide à la détection dans le contexte du contrôle d'aspect industriel, il est intéressant de remarquer que les anomalies les plus fréquemment rencontrées présentent très souvent un caractère non-périodique (coups, marques, rayures, etc.).

A la lumière de cette observation, Le Goïc justifie la pertinence de l'utilisation de la DMD pour caractériser ce type d'anomalies, car elle est construite sur une base de descripteurs périodiques. L'utilisation de ce paramétrage permet donc de caractériser l'ensemble des composantes périodiques d'une surface sur une large gamme de longueurs d'ondes. Elle permet également d'isoler les composantes non-périodiques des autres composantes, facilitant la caractérisation des anomalies d'aspect se distinguant par leur caractère non-périodique [Le Goïc2012].

Le Goïc introduit deux types d'approches pour l'aspect d'une surface :

- l'approche **géométrique**, qui fait le lien entre certains attributs géométriques (altitude, pentes et courbures⁵ de la surface et son apparence ;
- l'approche **lumière**, qui considère l'aspect de la surface du point de vue de son interaction avec son environnement lumineux.

La DMD peut être appliquée dans les deux cas. Dans le cadre de l'approche *géométrique*, elle permet de caractériser des variations dans la forme de la surface. Dans le cadre de l'approche *lumière*, elle permet de caractériser des variations de niveaux de gris, qui correspondent à des variations de la manière dont la surface réfléchit la lumière.

Partant du constat de pertinence de l'application de DMD à l'analyse de l'aspect visuel de surface, deux applications majeures ont été proposées dans les deux types d'approches précédemment cités.

Dans le cadre de l'approche géométrique, Le Goïc a présenté une technique de filtrage permettant de caractériser la surface à différents niveaux d'échelle. Il s'agit du *filtrage modal* [Goïc2011].

Dans le cadre de l'approche lumière, Le Goïc a introduit l'idée d'une nouvelle méthode d'approximation de la surface de réflectance, caractérisant la manière dont le flux lumineux reçu par une surface est réémis par celle-ci, basée sur la DMD [Le Goïc2012]. Cette méthode a ensuite été développée et mise en œuvre par Pitard [Pitard2017a].

Nous présentons ci-après les principes de ces deux développements, afin de permettre au lecteur de mieux comprendre l'intérêt qu'elle présente pour notre application à la description d'images, que nous décrivons par la suite.

2.3.4 Analyse multi-échelle par filtrage modal

L'opération de projection d'une surface dans la base des modes proposés permet donc d'obtenir une décomposition sous la forme d'une combinaison linéaire décrite par l'équation 2.3. Cette combinaison est représentée graphiquement grâce à un spectre modal, représenté en Figure 2.14. Le principe du *filtrage modal* repose sur la division de ce spectre en différentes parties, caractérisant les différents niveaux de variations surfaciques du point de vue fréquentiel. On retrouve ainsi les variations surfaciques de *forme*, d'*ondulation* et de *rugosité*, mais également une dernière partie contenant des composantes non-périodiques et du bruit de mesure. Ce dernier segment du spectre constitue un terme ε nommé *résidu*, qui permet donc de séparer les composantes périodiques et continues des composantes non-périodiques et non-continues de la surface. Le principe

5. L'altitude correspond à une topographie des hauteurs des points composant la surface. Les pentes correspondent aux normales en chaque point de la surface et peuvent être obtenues par dérivation de l'information en altitude. Les courbures sont quant à elles obtenues par dérivation de l'information en pentes.

de ce filtrage est illustré en Figure 2.15.

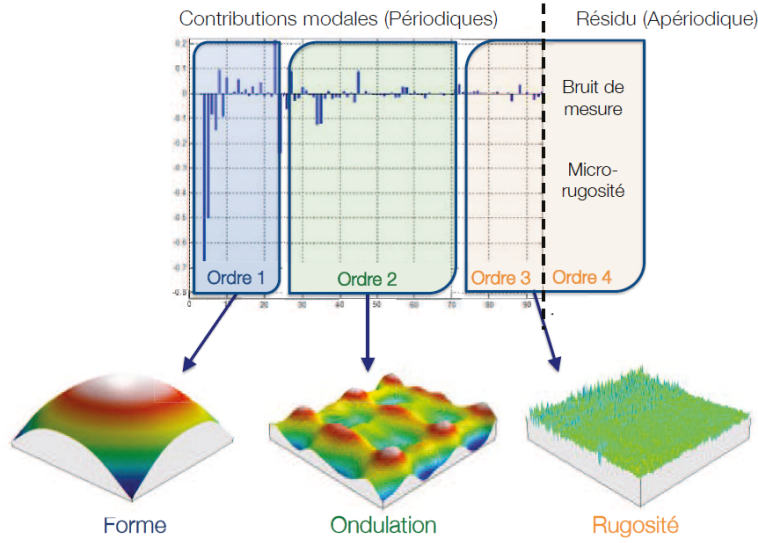


Figure 2.15 – Représentation de l'analyse multi-échelle d'un spectre modal, issue des travaux de [Le Goïc2012]

Les composantes de formes, ondulation et rugosité peuvent être obtenues en recomposant la partie du spectre modal correspondante. Cette opération est exprimée par les équations suivantes :

$$\begin{cases} formV = \sum_{i=1}^{N_f} \lambda_i \times Q_i \\ onduV = \sum_{i=N_f+1}^{N_o} \lambda_i \times Q_i \\ rugoV = \sum_{i=N_o+1}^N \lambda_i \times Q_i \end{cases} \quad (2.5)$$

avec N le nombre total de modes utilisés, N_f et N_o respectivement les seuils Forme/Ondulation et Ondulation/Rugosité. Ces deux seuils sont à choisir suivant la surface étudiée. Plus de détails quant au choix de ces seuils sont disponibles dans la thèse de Le Goïc [Le Goïc2012].

Le résidu $\varepsilon(N)$ de l'équation 2.3 peut alors être obtenu en soustrayant la surface reconstruite au travers des N modes de la DMD à la mesure ^{mes}V :

$$\varepsilon(N) = ^{mes}V - \sum_{i=1}^N \lambda_i \times Q_i \quad (2.6)$$

Un exemple d'application sur une surface métallique, issu de la thèse de Le Goïc, est exposé en Figure 2.16. Elle présente la reconstruction des surfaces résultats pour les composantes de forme et d'ondulation, ainsi que pour le résidu.

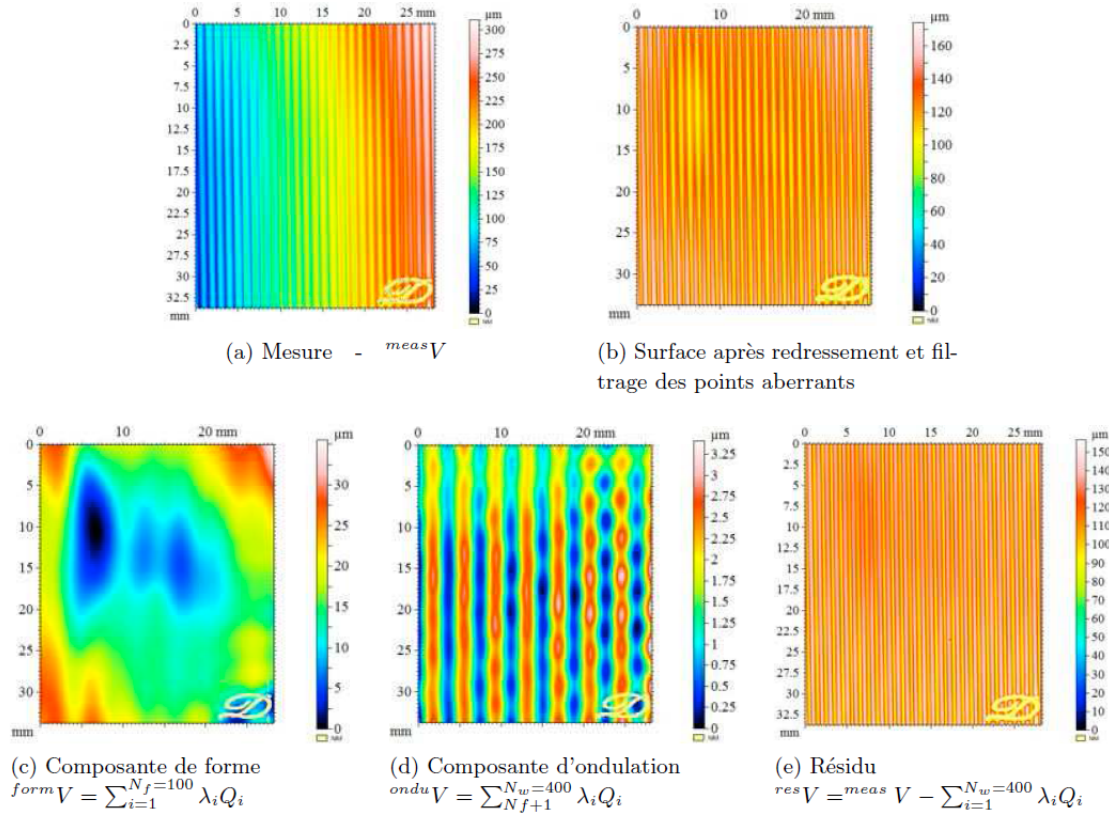


Figure 2.16 – Exemple d'analyse multi-échelle sur une mesure de surface métallique, issu de [Le Goïc2012]

A partir de cette analyse, il est possible de mettre en évidence et d'isoler certaines variations de forme pouvant expliquer des anomalies d'aspect, et de le caractériser géométriquement. Nous ne détaillons pas ces résultats ici mais nous invitons le lecteur à lire le travail de LeGoïc pour plus de précisions.

Nous souhaitons maintenant présenter les travaux menés sur l'application de la DMD à l'approche lumière de l'aspect des surfaces, avec les travaux menés sur l'approximation de la surface de luminance pour la reconstruction de surface RTI (*Reflectance Transformation Imaging*).

2.3.5 Approximation de la surface de luminance

La fonction, ou surface, de *luminance* est une modélisation de la lumière réfléchiée par une surface. Les techniques de reconstruction de rendu de surface RTI, pour *Reflectance Transformation Imaging*, utilisent la mesure de cette fonction afin d'enrichir l'information fournie pour des images classiques. En faisant varier les positions d'éclairages lors de l'acquisition d'images d'une surface, elle permet de décrire chaque pixel par un ensemble de valeurs, chaque valeur correspondant à une position précise d'éclairage. La reconstruction consiste alors en une interpolation, ou approximation, de l'ensemble continu des valeurs décrivant la luminance d'un pixel (fonction de luminance du pixel). Ces techniques permettent d'obtenir une surface virtuelle après reconstruction, présentant une information enrichie pour mettre mieux en valeur les interactions locales entre surface et lumière.

L'application de la DMD à la reconstruction RTI a permis d'améliorer les performances des techniques déjà existantes, telles que les PTM (*Polynomial Texture Mapping*) et les HSH (*Hemispherical Harmonic*) [Pitard2017a]. Au lieu de décomposer la mesure géométrique d'une surface

dans une base modale comme précédemment, ce sont les surfaces de luminance qui sont décomposées à partir des mesures multi-éclairages.

La Figure 2.17 présente un comparatif visuel des reconstructions de la surface de luminance d'un même pixel par les trois méthodes PTM, HSH et DMD. La reconstruction utilisant la DMD utilise les 49 premiers modes de la décomposition. Il est possible de constater que les variations de luminance sont mieux approximées avec la DMD, notamment pour les comportements spéculaires⁶. Pour compléter cette analyse, la Figure 2.18, présente le comparatif des rendus d'une même zone d'intérêt d'un anneau métallique sous deux angles d'éclairage précis.

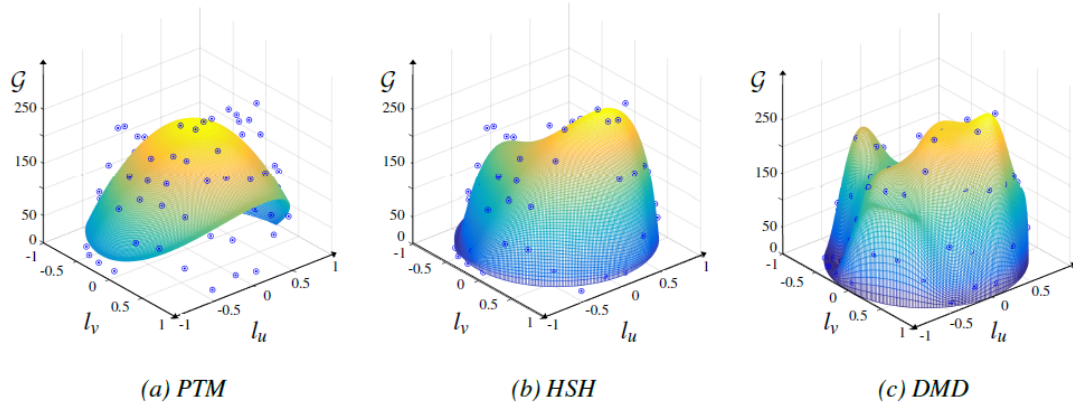


Figure 2.17 – Surfaces de réflectance approximées, pour un même pixel par les méthodes PTM, HSH et DMD, issues de [Pitard2016]

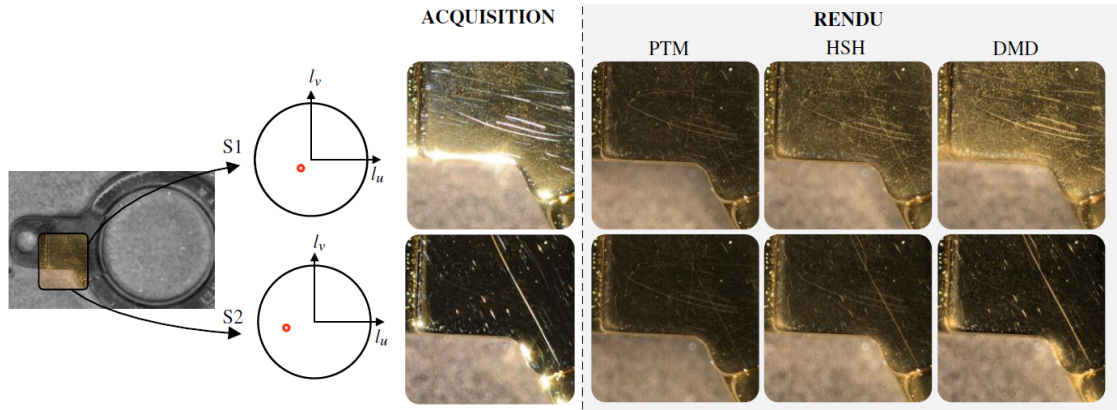


Figure 2.18 – Exemple de rendu avec les méthodes PTM, HSH et DMD, pour deux directions d'éclairage S1 et S2, issu de [Pitard2016]

De manière quantitative, la performance de la DMD a été comparée avec des critères classiques de comparaison d'images (SSIM et PSNR) et également grâce à un des paramètres statistiques de forme (kurtosis et skewness) calculés sur les distributions des luminances de chaque pixel des reconstructions RTI. Les résultats obtenus montrent que la DMD présente des performances équivalentes pour l'approximation des luminances diffuses, mais qu'elle permet d'obtenir des résultats plus fidèles pour le comportement de luminance spéculaire local [Pitard2016].

6. Un comportement spéculaire décrit une réflexion importante de la lumière dans une direction privilégiée, se traduisant par des pics sur la surface de luminance.

A partir de la modélisation de la luminance par la DMD, Pitard a également introduit une approche de recherche de similarité entre surfaces. Celle-ci se base sur l'utilisation des coefficients modaux comme d'une signature de la surface étudiée.

L'apport de nos travaux dans ce chapitre se fonde sur l'extension de l'utilisation du spectre modal comme d'une signature de l'information contenue dans une image, utilisable comme un *descripteur* pour nourrir par la suite un algorithme de classification automatique.

2.3.6 Description d'images via la DMD

Les travaux précédents ont montré la pertinence de l'approche modale pour tout d'abord paramétrer géométriquement une surface, mais également pour approximer les surfaces de luminances des pixels. Ces résultats indiquent que la DMD d'une image géométrique ou optique d'une surface contient des informations caractéristiques pour décrire cette dernière.

Partant de ce constat, nous souhaitons étendre l'utilisation de cette approche pour extraire des *descripteurs* d'image. Comme présenté précédemment, le rôle des descripteurs est de caractériser l'information de l'image tout en la compressant en une représentation plus simple. Le calcul du spectre modal d'une image répond à cette définition et nous proposons donc de l'utiliser dans ce sens. Pour faire référence aux familles de méthodes présentées précédemment, la DMD s'inscrit dans la catégorie des *approches par filtrage ou transformation*. En effet, elle consiste bien en une décomposition de l'image dans une base d'éléments simples.

Nous proposons deux approches pour l'utilisation de la DMD pour la description d'images :

- l'approche **coefficients modaux**, qui correspond à l'utilisation directe des coefficients issus de la décomposition de l'image dans la base modale. Il est à noter que ces coefficients peuvent être nombreux et peuvent présenter une information non pertinente ou redondante les uns par rapport aux autres. Il est donc souvent nécessaire de réaliser un tri de ces derniers pour ne garder que les plus pertinents. Le Chapitre 3 de ce manuscrit donne des clefs au lecteur pour réaliser cette sélection.
- l'approche **paramètres statistiques modaux**, qui correspond à l'utilisation de paramètres statistiques calculés sur la distribution des valeurs du spectre modale. En effet, au lieu d'utiliser les valeurs particulières de chaque mode, nous proposons également d'utiliser des paramètres caractérisant la distribution des valeurs du spectre modal. Nous avons pour cela utilisé des paramètres statistiques du premier ordre tel que la médiane, le minimum, le maximum, le kurtosis ou encore le skewness.

Afin d'illustrer la pertinence de notre proposition, nous avons évalué les performances de l'utilisation de coefficients modaux sur un problème de classification simple, en comparaison de l'utilisation des descripteurs d'Haralick, dont l'usage est répandu pour ce genre d'applications. Notre objectif ici est d'évaluer la capacité des coefficients modaux à décrire l'information d'une image.

Conditions expérimentales

Lot de données

Pour mener cette étude, nous avons utilisé des plaquettes charmillles⁷, étalons de rugosité en plastique, dont un exemple est visible en Figure 2.19. Nous avons séparé les plaquettes en six *patches* de rugosités différentes, illustrés en Figure 2.20. Les images sont de dimensions 800×500 pixels, et ont toutes été acquises sous le même angle d'éclairage.



Figure 2.19 – Plaquette charmille présentant des étalons de différentes rugosités

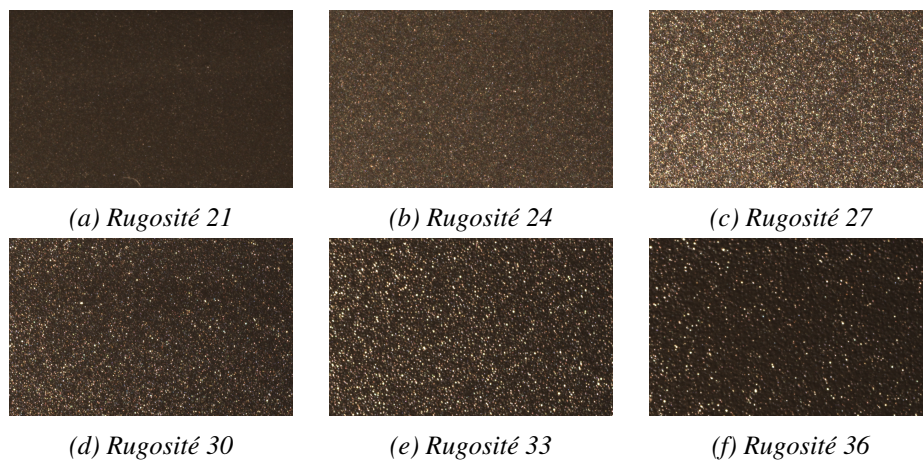


Figure 2.20 – Patches de rugosités utilisés pour la classification à 6 classes

Un lot de 50 plaquettes a été utilisé, permettant ainsi d'obtenir 50 éléments pour chaque patch de rugosité. Ce lot de données a ensuite été séparé en 30 éléments d'apprentissage et 20 éléments de test pour chaque classe de rugosité. Les données d'apprentissage ont servi à entraîner un algorithme de classification, alors que les données de test ont été utilisées afin d'évaluer les performances de la classification.

7. Cette appellation vient de l'entreprise Charmilles, qui a développé une échelle de rugosité. Cette dernière est référencée dans les normes et est donc d'usage commun. Pour plus d'information à ce sujet, nous invitons le lecteur à consulter le site de [Metals-Industry](#).

Algorithmes de classification

Deux types d'algorithmes de classification ont été utilisés : une *machine à vecteur de support* (SVM) One-Versus-Rest, et un *arbre de décision*. Le lecteur peut retrouver les principes fonctionnement de ces techniques dans la Chapitre 4 de ce mémoire, dédié aux algorithmes de classification.

Descripteurs

Pour l'approche *coefficients modaux*, nous avons utilisé une décomposition dans une base comportant 75 modes, permettant ainsi d'obtenir 75 descripteurs pour une image.

Pour l'approche *paramètres statistiques modaux*, nous avons extrait la médiane, le minimum, le maximum, le kurtosis et le skewness de la distribution des valeurs des coefficients modaux. Les 5 valeurs obtenues ont été utilisées pour la classification.

Afin d'avoir un élément de référence auquel nous rapporter, nous avons utilisé les 14 premiers coefficients d'Haralicks extraits des matrices de cooccurrences des images (Section 2.2.2.2).

L'étude menée se déroule en 3 phases :

- Classification 3 classes (rugosités 21, 30 et 36) avec les *coefficients modaux*,
- Classification 6 classes (rugosités 21, 24, 27, 30 et 36) avec les *coefficients modaux*,
- Classification 6 classes (rugosités 21, 24, 27, 30 et 36) avec les *paramètres statistiques modaux*.

Classification à 3 classes - coefficients modaux

Pour cette première phase, le choix a été fait de n'utiliser que 3 niveaux de rugosités, de niveaux extrêmes (21 et 36) et intermédiaire (30), afin d'éprouver l'utilisation des coefficients modaux sur un cas très simple dans un premier temps. Les résultats de la classification des éléments de test sont présentés dans le Tableau 2.2 pour les coefficients modaux, et le Tableau 2.1 pour les paramètres d'Haralick. Ils sont présentés sous forme de tableau à double entrée présentant le nombre d'éléments bien classés pour chaque classe de rugosité sur la diagonale (haut gauche / bas droit), et le nombre d'éléments mal classés dans les autres cases. Le classifieur utilisé est le SVM qui a apporté les meilleurs résultats dans ce cas.

Prédit \ Théo	Rugo21	Rugo30	Rugo36
Rugo21	20		
Rugo30		20	
Rugo36			20

Tableau 2.1 – Résultat de la classification de 60 pièces de test réparties en 3 classes et des descripteurs d'Haralick

Prédit \ Théo	Rugo21	Rugo30	Rugo36
Rugo21	19		1
Rugo30		20	
Rugo36	1		19

Tableau 2.2 – Résultat de la classification de 60 pièces de test réparties en 3 classes et des descripteurs coefficients modaux

La première observation réalisée est que les paramètres d'Haralick permettent d'obtenir une performance de 100% de bonne classification pour les différents éléments de test. Ces attributs sont reconnus pour être efficace dans la caractérisation de textures d'images. Il n'est donc pas surprenant qu'ils permettent de bien décrire et classer les différentes rugosités, qui présentent des textures différentes. Ces résultats se présentent donc comme une référence justifiée.

La seconde observation porte sur les performances de la classification via les coefficients modaux. Ils permettent quant à eux d'obtenir une performance de 97% de bonne classification pour les éléments de test (2 erreurs sur 60). Ces résultats sont comparables à ceux obtenus avec les paramètres d'Haralick.

A la suite de cette première phase portant sur seulement 3 classes très différenciées, nous avons introduit 3 classes supplémentaires présentant des rugosités intermédiaires (24, 27 et 33).

Classification à 6 classes - coefficients modaux

Cette deuxième phase porte donc sur la classification de 6 niveaux de rugosité (21, 24, 27, 30, 33 et 36), toujours avec l'utilisation des coefficient modaux. Les résultats pour la classification à partir des paramètres d'Haralick et des coefficients modaux sont quant à eux exposés dans le Tableau 2.3, selon le même principe que précédemment. Le classifieur utilisé est le SVM qui a également apporté les meilleurs résultats dans ce cas.

Prédit \ Théo	Rugo21	Rugo24	Rugo27	Rugo30	Rugo33	Rugo36
Rugo21	20 / 19					1
Rugo24		20 / 20				
Rugo27			20 / 20			
Rugo30				20 / 12	4	
Rugo33				8	20 / 16	
Rugo36	1					20 / 19

Tableau 2.3 – Résultat de la classification de 120 pièces de test réparties en 6 classes, avec les descripteurs d'Haralick en bleu, et les coefficients modaux en noir.

Les performances de la classification exploitant les coefficients modaux sont ici de 88% de bonnes décisions. Les performances sont donc plus faibles que pour la classification sur 3 classes

et l'algorithme confond particulièrement les rugosités 30 et 33. Néanmoins, les résultats restent très bons pour les autres classes. Notre approche est donc dans ce cas moins performante que les paramètres d'Haralick, qui continuent à permettre de parfaitement discriminer les différentes classes.

Pour la dernière phase de cette étude, nous avons utilisé cette fois-ci les paramètres statistiques calculés sur la distribution des coefficients, toujours avec 6 classes de rugosité.

Classification à 6 classes - paramètres statistiques du spectre

Cette dernière phase porte donc sur l'utilisation de 5 paramètres statistiques (médiane, minimum, maximum, kurtosis et skewness) calculés sur la distribution des valeurs des coefficients modaux, dans les mêmes conditions de classification que la phase précédente. Néanmoins, cette fois-ci, les résultats de l'arbre décisionnel étaient meilleurs que pour le SVM. Ils sont présentés dans le Tableau 2.4.

Théo \ Prédit	Rugo21	Rugo24	Rugo27	Rugo30	Rugo33	Rugo36
Rugo21	20 / 20					3
Rugo24		20 / 20			1	
Rugo27			20 / 20	1		
Rugo30				20 / 19		
Rugo33					20 / 19	
Rugo36						20 / 17

Tableau 2.4 – Résultat de la classification de 120 pièces de test réparties en 6 classes avec les descripteurs d'Haralick en bleu, et les paramètres statistiques modaux en noir (paramètres statistiques calculés sur le spectre des coefficients)

Pour cette dernière phase, les performances de la classification s'élèvent à 96% de bonnes décisions, améliorant les performances obtenues précédemment avec les coefficients modaux directs. On se rapproche ainsi des performances obtenues avec les paramètres d'Haralick.

Discussion

Les résultats obtenus avec l'utilisation de la DMD pour décrire une image dans un objectif de classification montrent que cette approche permet de caractériser l'information contenue dans une image. En effet, les résultats obtenus sont globalement bons, comparables avec ceux obtenus avec les paramètres d'Haralick, même si ils semblent néanmoins un peu moins performants sur ce cas d'application.

La Figure 2.21 montre l'évolution de différents paramètres utilisés pour décrire les patches de rugosité, afin de mieux comprendre et illustrer les résultats présentés précédemment. Les graphiques exposent les valeurs des paramètres pour chaque donnée d'entraînement, regroupées par classe de rugosité.

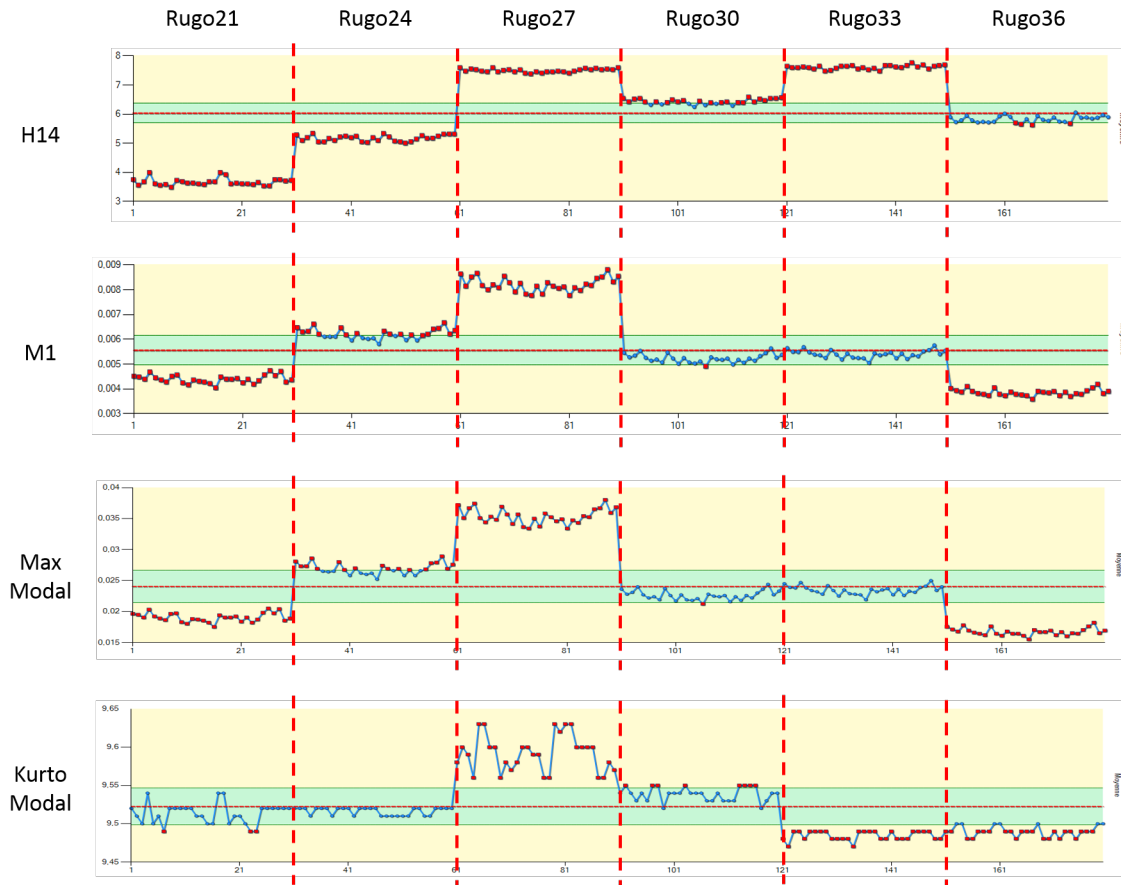


Figure 2.21 – Evolution de certains des paramètres utilisés pour la classification des rugosités. De haut en bas : 14ème paramètre d'Haralick, 1er coefficient modal, maximum et kurtosis de la distribution des coefficients modaux.

Ces graphiques permettent de mieux se rendre compte de l'information contenue dans les descripteurs extraits. Nous exposons des exemples de paramètres présentant le plus de pertinence pour l'application de classification souhaitée. Nous définissons cette pertinence par une réponse différente du descripteur pour les différents niveaux de rugosité.

Dans ce contexte, les très bonnes performances de la classification avec les *paramètres d'Haralick* se justifient par l'observation de l'évolution des paramètres extraits. En effet, à l'image du paramètre H14 présenté en Figure 2.21, les paramètres répondent de manière distincte pour chaque type de rugosité. Les autres paramètres présentent globalement des évolutions similaires, ce qui explique la bonne séparation des classes.

Pour ce qui s'agit des paramètres issus directement des *coefficients modaux*, certains d'entre-eux présentent également une évolution pertinente, à l'image du paramètre M1 présenté en Figure 2.21. Néanmoins, aucun paramètre parmi ceux extraits ne semble proposer une séparation assez distincte entre les réponses aux images de la rugosité 30 et à celles de la rugosité 33. Cela peut expliquer les erreurs de classification observées en phase 2 de l'étude.

Enfin, l'extraction des paramètres statistiques du spectre modal permet d'obtenir des paramètres également pertinents, à l'image du maximum et du kurtosis présentés en Figure 2.21. La différence avec l'extraction des coefficients directs est que certains paramètres extraits permettent de mieux séparer les classes 30 et 33, comme le kurtosis ici. La combinaison entre le maximum qui décrit bien la séparation entre les classes 21, 24, 27, 30/33 et 36 et le kurtosis qui décrit bien la séparation entre les classes 30 et 33 permet ainsi d'obtenir de meilleurs résultats globaux de classification.

A travers cette analyse, nous mettons donc en avant le fait que les descripteurs extraits via

l'approche DMD que nous proposons contiennent une information pertinente pour réaliser des tâches de classification.

2.4 Travaux applicatifs utilisant la description d'images

Si l'effort principal de mon travail de thèse est focalisé sur le développement d'une application d'inspection visuelle automatisée en industrie, j'ai également pu apporter ma contribution en terme de description d'images à différents autres projets de recherche.

Pour clore ce chapitre, deux travaux applicatifs de la description d'images, notamment avec la DMD, sont donc présentés. Ils illustrent certaines pistes de travail autres que l'application à l'inspection automatisée de surfaces, qui est elle développée dans le dernier chapitre de ce mémoire. Ces travaux ont été menés avec la volonté de créer des collaborations avec d'autres doctorants du laboratoire SYMME et ainsi nouer des liens entre différents projets portant sur des problématiques de qualité industrielle.

La première étude présentée porte sur l'extension de l'utilisation de la DMD à la mesure de performance de prédiction d'un réseau de Deep Learning. Elle a été menée en collaboration avec Pierre Nagorny, dont les travaux de thèse portent sur le pilotage en temps réel de presses d'injection plastique.

La seconde étude porte quant à elle sur l'utilisation de la description d'images pour caractériser des propriétés haptiques de surfaces. Elle a été menée en collaboration avec Bruno Albert, dont les travaux de thèse portent sur la formalisation du contrôle qualité haptique en industrie, et un groupe d'étudiants de Polytech Annecy-Chambéry, menant leur projet de fin d'études au laboratoire.

2.4.1 La DMD comme un indicateur de performance en Deep Learning

Au-delà de son utilisation pour décrire une image, nous avons également appliqué la DMD comme un indicateur de performance pour analyser les prédictions d'un réseau de Deep Learning. Ce travail a fait l'objet d'une communication en congrès [Nagorny2018].

Contexte de l'étude

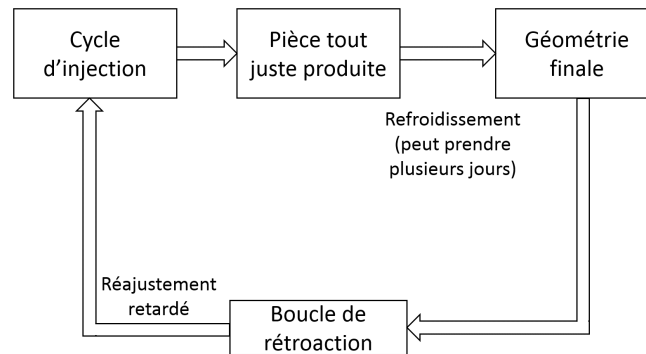
Cette étude s'inscrit dans le domaine du pilotage des machines d'injection plastique. Les travaux de Pierre Nagorny portent sur la mise en œuvre d'un pilotage automatique des paramètres d'injection, en fonction de mesures effectuées sur les pièces en sortie de moule. L'objectif est de parvenir à une production se réajustant d'elle-même en temps réel pour toujours produire des pièces conformes.

Or, lors du procédé d'injection plastique, les pièces produites mettent un certains temps (quelques heures à quelques jours) à refroidir et à prendre leur forme finale. Étant donné que l'objectif est de piloter la presse en temps réel, il est impossible d'utiliser des mesures effectuées après refroidissement. Il est alors indispensable de concevoir un modèle pour prédire la géométrie finale d'une pièce via des mesures en sortie de moule, comme illustré en Figure 2.22.

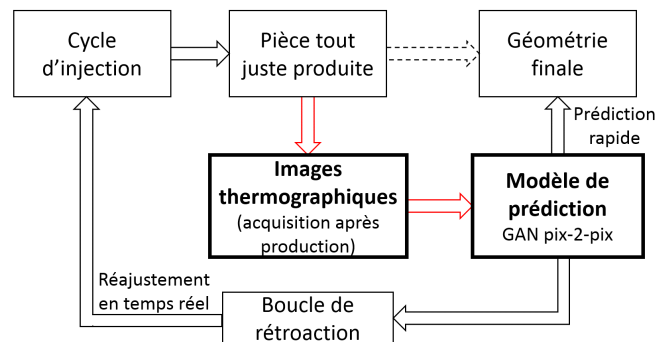
L'objectif de cette étude est donc de mettre au point un modèle de prédiction de la géométrie finale de pièces plastiques à partir d'images thermographiques acquises en sortie de moule.

Nous avons utilisé un réseau de Deep Learning capable de réaliser cette tâche à partir de peu

d'éléments d'apprentissage.



(a) Ajustement classique



(b) Ajustement basé sur un modèle de prédiction

Figure 2.22 – Représentation du processus d'ajustement des paramètres du cycle d'injection plastique

Si Pierre Nagorny a apporté son expertise sur les réseaux de Deep Learning pour développer la génération des images, mon apport pour cette étude s'est plutôt porté sur l'utilisation de la DMD pour tester la qualité de prédiction du réseau. La suite de cette section s'applique à développer succinctement cet apport, ainsi que de présenter rapidement les résultats de l'étude.

Évaluer la qualité de prédiction avec la DMD

Afin d'évaluer la performance prédictive du réseau, il est possible de comparer les géométries générées aux mesures réelles. Pour cela, un lot de 14 mesures tests a été gardé en dehors de l'apprentissage du réseau, afin de pouvoir réaliser cette validation.

Parmi les indicateurs classiques de performance pour la génération d'images, on retrouve souvent le PSNR (Peak Signal to Noise Ratio), le SSIM (Structural Similarity) ou encore certaines mesures de distances comme la distance cosinus ou la distance de corrélation. Or, ces indicateurs sont conçus pour mesurer des similarités pixel à pixel ou de structure entre des images, du point de vue visuel.

Dans notre cas, nous ne générons pas d'images classiques mais bien des géométries de surfaces. Il semble donc nécessaire d'introduire un nouvel indicateur permettant de mesurer la similarité géométrique des surfaces générées. Pour cela, nous proposons d'utiliser la DMD en comparant les spectres modaux des géométries générées et mesurées, afin de mesurer leur niveau de similarité et ainsi d'évaluer la capacité du réseau à récupérer les caractéristiques géométriques des surfaces.

Nous avons donc comparé les spectres des 14 pièces tests, sur les géométries mesurées et prédites. Deux comparaisons de spectres sont présentées en Figure 2.23.

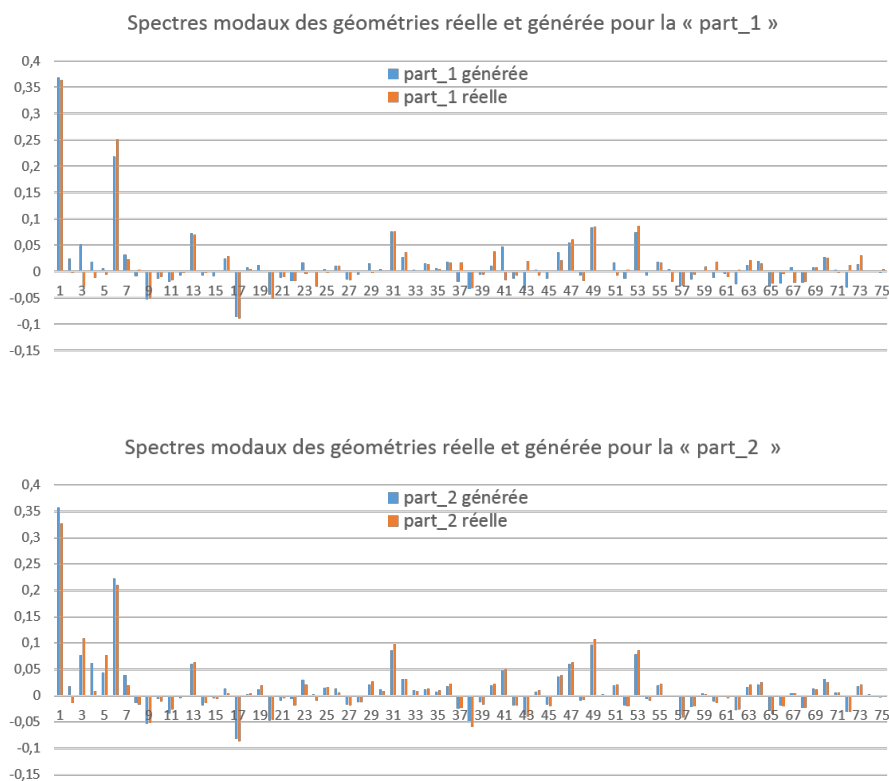


Figure 2.23 – Comparaison des spectres modaux des géométries réelles et générées pour deux pièces du lot de test

Pièce	Similarité entre les spectres des géométries réelles et générées	
	Distance cosinus	Coefficient de corrélation R^2
Part 1	0.94	0.88
Part 2	0.98	0.96
Médiane 14 pièces	0.99	0.98
Écart-type 14 pièces	0.04	0.08

Tableau 2.5 – Similarité entre les spectres

Dans un premier temps, il est possible de constater que les spectres semblent très proches visuellement, et plus spécialement pour les modes présentant de fortes amplitudes, donc qui participent le plus à la composition de la surface.

Dans un second temps et afin d'obtenir une mesure quantitative de cette similarité, nous avons comparé les spectres grâce à deux indicateurs de similarités couramment utilisés pour confronter des vecteurs de données : la distance cosinus et le coefficient de corrélation. Les résultats sont présentés en Tableau 2.5.

Ces résultats permettent de constater une très bonne similarité entre les spectres sur l'ensemble des pièces de test. Ces similarités démontrent la capacité du réseau à retrouver les caractéristiques géométriques des surfaces pour ce cas d'application particulier.

Au delà des résultats prometteurs dans le domaine du pilotage des machines d'injection, ces travaux ouvrent différentes perspectives quant à l'utilisation de la DMD dans ce type d'application. Nous proposons de les exposer dans la suite de cette section.

Perspectives concernant l'utilisation de la DMD comme indicateur de performance

Dans cette étude, nous avons donc introduit la DMD comme un indicateur innovant pour évaluer la capacité d'un réseau de neurone à générer des surfaces en préservant leurs caractéristiques géométriques.

Pour aller plus loin dans l'utilisation de cet outil, nous proposons d'analyser les spectres non plus globalement, mais plutôt mode par mode. Par analyse globale, nous entendons étude de similarité entre deux spectres, en considérant tous les modes générés. Si cette analyse permet de donner une idée générale de la similarité entre deux surfaces, une étude "mode à mode" semble également intéressante pour identifier quels modes sont les mieux prédits par le réseau, et donc quels types de formes et déformations sont les mieux générés.

Enfin, cet outil pourrait également servir à mieux comprendre le fonctionnement de la génération des images à travers le réseau de neurones. En effet, les réseaux de neurones, et plus spécialement de Deep Learning, sont considérés comme des "boîtes noires" car une fois entraînés, il est très difficile d'interpréter comment ils parviennent à générer une certaine sortie. De nombreux travaux portent à l'heure actuelle sur ces problématiques dans le domaine de l'intelligence artificielle notamment [[Gunning2017](#)].

La génération et l'analyse des spectres modaux des images des géométries à chaque couche du réseau de neurones pourraient éventuellement apporter des informations sur le processus de construction de ces dernières. En examinant l'évolution du spectre et des différents modes à chaque couche du réseau, il est peut être en effet possible de mieux comprendre comment fonctionne le réseau. Y a-t-il une logique compréhensible dans la construction successive des géométries ? Les couches du réseau se spécialisent-elles dans la reconstruction de modes spécifiques, ou participent-elles à tous les modes en même temps ? Telles sont les questions auxquelles nous aimerions par exemple pouvoir répondre.

2.4.2 Description d'images pour prédire des caractéristiques haptiques

Pour cette étude, nous proposons une utilisation de la description d'image dans un objectif de caractérisation de surfaces du point de vue des sensations haptiques. Ce travail a également fait l'objet d'une communication en congrès [[Lacombe2017](#)].

Contexte de l'étude

Nous avons exposé l'importance des contrôles de la qualité visuelle des produits industriels dans le premier chapitre de ce mémoire. Or, il existe d'autres types de contrôles dits *sensoriels* mis en œuvre dans le milieu industriel. C'est le cas du contrôle *haptique* lié au sens du toucher. Bien que le contrôle de la qualité visuelle d'un produit ait été déjà relativement bien étudié et instrumenté, la qualité associée au sens du toucher pose encore de nombreux problèmes dans sa mise en place. De plus, dans certaines applications, le contact direct avec les produits n'est pas possible ou entraînerait des dégradations. Ce travail exploratoire vise à étudier des pistes de travail pour la substitution du contrôle au toucher par un contrôle visuel sans contact.

Dans un premier temps, nous avons décidé de concentrer cette étude sur la sensation d'*adhérence*, qui peut être définie comme la sensation d'étirement ou de retenue de la peau lors du contact avec la surface d'un objet.

La problématique a été décomposée en deux aspects : la capacité de la perception visuelle humaine à prévoir une sensation haptique, et la possibilité de caractérisation instrumentée via par exemple des mesures optiques.

Dans un premier temps, l'étude s'est donc focalisée sur la performance visuelle humaine afin de mesurer la sensation d'adhérence. Puis dans un second temps, il a été question d'étudier une éventuelle relation entre une mesure optique instrumentale sans contact et une mesure de contact liée à la sensation d'adhérence.

Les résultats de la première partie de l'étude sont brièvement présentés ci-après, puis nous nous attardons plus longuement sur la seconde phase dont les résultats nous intéressent plus dans le cadre de ce chapitre.

Performance de la perception visuelle humaine

Le protocole expérimental de cette première phase a visé à faire passer des tests visuels et haptiques à 15 participants sur 12 échantillons de matières, présentés dans la Figure 2.25. Il s'agit de films polymères collés sur des plaques rigides, présentant des degrés d'intensité de la sensation d'adhérence différents. La Figure 2.25 rassemble quelques caractéristiques de ces plaques, définies de manière empirique par consensus dans l'équipe menant l'étude.

Les échantillons ont été couplés par paires et les participants ont eu pour tâche de comparer l'adhérence de chaque échantillon avec son binôme visuellement et haptiquement. Pour chaque paire, une note de 1 était attribuée si l'échantillon 1 était considéré comme le plus adhérent, -1 si l'échantillon 2 était considéré comme le plus adhérent et 0 si les deux étaient perçus comme similaires. Les résultats de cette expérience moyennés sur les 15 participants sont présentés en Figure 2.24. On observe qu'en général, les tests visuels ne donnent pas de résultats assez significatifs pour conclure sur une différence entre les échantillons proposés, contrairement aux tests haptiques pour lesquels les participants sont parvenus à identifier la différence d'adhérence de manière plus significative.

Ces premiers résultats permettent de se rendre compte que le sens visuel humain ne semble pas adapté pour différencier ces échantillons par rapport à la perception de la sensation d'adhérence. Naturellement, le sens du toucher est plus approprié pour cela.

N° paire			Visuel			Haptique		
	Echan tillon 1	Echan tillon 2	Ratio moyen du plus adhérent	Ecart type	Echantillon le plus adhérent (*=significatif)	Ratio moyen du plus adhérent	Ecart type	Echantillon le plus adhérent (*=significatif)
1	7	9	-0,33	0,49	(9)	0,60	0,41	7 *
2	8	10	0,07	0,52	(8)	0,60	0,41	8 *
3	11	13	-0,73	0,35	13 *	-0,87	0,26	13 *
4	6	26	-0,20	0,51	(26)	0,47	0,46	6 *
5	14	19	-0,07	0,52	(19)	-0,60	0,41	19 *
6	12	31	0,20	0,51	(12)	-0,33	0,49	(31)

Figure 2.24 – Résultats des performances de perception de la sensation d'adhérence à l'œil et au toucher

Une seconde expérience nous a également permis de nous rendre compte que la perception visuelle peut même tromper la perception haptique quand les sens de la vue et du toucher sont utilisés de manière conjointe.

Ces observations nous ont permis de préconiser qu'il serait inapproprié, voir contre-indiqué d'utiliser le sens de la vue à des fins de contrôle haptique, pour ce qui est de la sensation d'adhérence.

N° plaque	Visuel				Haptique			Image
	Couleur	Aspect matière	Brillance /5	Uniformité /3	Relief /3	Dureté /3	Uniformité /3	
6	Noire	Feutré	1	3	2	2	3	
7	Noire	Plastique	4	2	1	1	2	
8	Noire	Plastique	3	2	1	3	2	
9	Anthracite	Feutré	2	2	2	1	2	
10	Anthracite	Plastique	2	1	1	2	3	
11	Anthracite	Plastique	5	2	1	3	3	
12	Noire	Plastique	5	2	3	3	2	
13	Anthracite	Ardoise	2	2	2	3	3	
14	Beige	Bois	2	2	1	3	3	
19	Beige	Bois	1	2	3	3	3	
26	Noire	Feutré	1	3	2	2	3	
31	Grise	Ardoise	1	2	1	3	3	

Figure 2.25 – Caractéristiques et illustrations des échantillons utilisés

Si l'humain ne semble pas performant pour ce type de prédiction, il semble ensuite intéressant d'étudier le potentiel de mesures instrumentées pour réaliser cette tâche.

Étude de corrélation entre instrumentation optique et perception haptique

Afin d'étudier le potentiel de prédiction de la sensation d'adhérence par des mesures optiques sans contact, nous avons tout d'abord cherché à caractériser la sensation d'adhérence à des mesures instrumentales. Ensuite, nous avons utilisé une instrumentation sans contact afin d'extraire

des caractéristiques que nous avons mises en regard des mesures précédentes pour étudier une éventuelle corrélation.

Nous proposons de détailler ces différentes étapes dans la suite de cette section.

Instrumentation avec contact : mesure d'angle de friction

Afin de caractériser l'adhérence des surfaces étudiées, l'angle de friction de chaque échantillon a été mesuré. Cette mesure a été réalisée en plaçant les pièces sur un plan et d'incliner celui-ci jusqu'à ce que l'échantillon se mette à glisser. Cette mesure simple a été effectuée sur chaque échantillon, en utilisant un plan incliné en plastique. L'erreur de mesure a été estimée à plus ou moins 3 degrés.

Dans le but de valider la relation entre ces mesures et la perception d'adhérence humaine, les valeurs d'angle de frottement trouvées ont été comparées aux résultats de la comparaison par paires présentés précédemment.

Les résultats de cette comparaison sont affichés en Figure 2.26. La colonne de gauche indique quel échantillon était perçu comme le plus adhérent pour chaque paire lors du test haptique. La colonne de droite indique la valeur de la différence d'angle entre le premier échantillon et le deuxième. Si la différence est positive, cela indique que la friction du premier angle est plus forte que celle du second, et inversement si elle est négative.

Ces résultats permettent de conclure que les mesures d'angles correspondent globalement bien aux comparaisons humaines. En effet, sauf pour une paire (11-13), toutes les mesures d'angles correspondent aux perceptions exprimées. Cela nous a donc conforté dans le fait d'utiliser ces mesures comme indicateur de la perception d'adhérence.

Résultats de comparaison haptique en termes d'adhérence	Différence d'angle (degrés)
Ech 7 > Ech 9	41
Ech 8 > Ech 10	16
Ech 11 < Ech 13	7
Ech 6 > Ech 26	8
Ech 14 < Ech 19	-6
Ech 12 < Ech 31	-21

Figure 2.26 – Comparaison des mesures d'angle de frottement aux résultats de perception haptique

Instrumentation sans contact : description d'images par descripteurs d'Haralick

Pour ce qui s'agit de la mesure optique sans contact, un dôme multi-éclairages a été utilisé. Il permet d'obtenir 96 images depuis 96 angles d'éclairages répartis à la surface du dôme, la caméra se trouvant au zénith de celui-ci, et l'objet au centre.

Nous avons dans un premier temps décidé d'utiliser 6 descripteurs d'Haralick pour décrire les images obtenues, issues de calculs de matrices de concurrence : le contraste, la dissimilarité, l'homogénéité, l'énergie, la corrélation et le second moment angulaire. Pour chaque descripteur et chaque pièce, 96 valeurs sont donc obtenues (une pour chaque images acquise). Afin de réduire le nombre de valeurs, 5 paramètres statistiques sont retenus pour décrire les distributions de 96 valeurs : les percentiles 5% et 95%, la médiane, le kurtosis et le skewness. Au final, chaque pièce est donc décrite par 30 paramètres (6 descripteurs d'Haralick x 5 paramètres statistiques).

L'objectif de la suite de l'étude a été de chercher s'il est possible d'établir des corrélations entre certains de ces paramètres issus des images et les mesures d'angles de friction précédentes.

Étude de corrélation

Afin de tenter d'établir une relation entre les valeurs des angles mesurés et des paramètres issus de la description des images, une régression multi-linéaire a été menée sur ces données. Elle a permis de dégager un modèle multi-linéaire reliant 5 des paramètres extraits aux valeurs d'angles. La relation entre les valeurs d'angles mesurés et les valeurs prédites par le modèle est illustrée en Figure 2.27 et présente un coefficient de coordination supérieur à 90%.

Conclusion et perspectives

Ces résultats permettent de montrer que dans ces *conditions particulières*, il est possible de modéliser de manière satisfaisante une relation multi-linéaire entre une mesure avec contact liée à l'adhérence de films polymères et de mesures optiques sans contact de ces mêmes surfaces. N'ayant pas plus d'échantillons à notre disposition que ceux utilisés pour calculer le modèle, aucune étude plus poussée n'a été menée quant à la capacité de prédiction pour des mesures additionnelles. Ceci constitue une des perspectives de travail pour la suite de cette exploration. Une autre perspective réside dans les outils utilisés pour modéliser la relation entre les données. Si nous avons choisi d'utiliser une relation multi-linéaire, d'autres méthodes existent comme la régression par réseaux de neurones qui peut être utilisée pour modéliser des relations plus compliquées. Cette technique a d'ailleurs été utilisée par Zhang, Dana et Nishino pour modéliser une relation entre une mesure de friction et une mesure de réflectance pour diverses surfaces [Zhang2016]. Enfin, l'utilisation d'autres types de descripteurs pour décrire les images ou l'exploitation d'autres types d'images ouvrent de nombreuses possibilités quant à la modélisation de corrélations. La combinaison entre les très nombreuses données qu'il est possible d'obtenir à partir de la description d'images et les différentes méthodes de modélisation existantes ouvre de vastes perspectives quant à la poursuite de ce type de travaux.

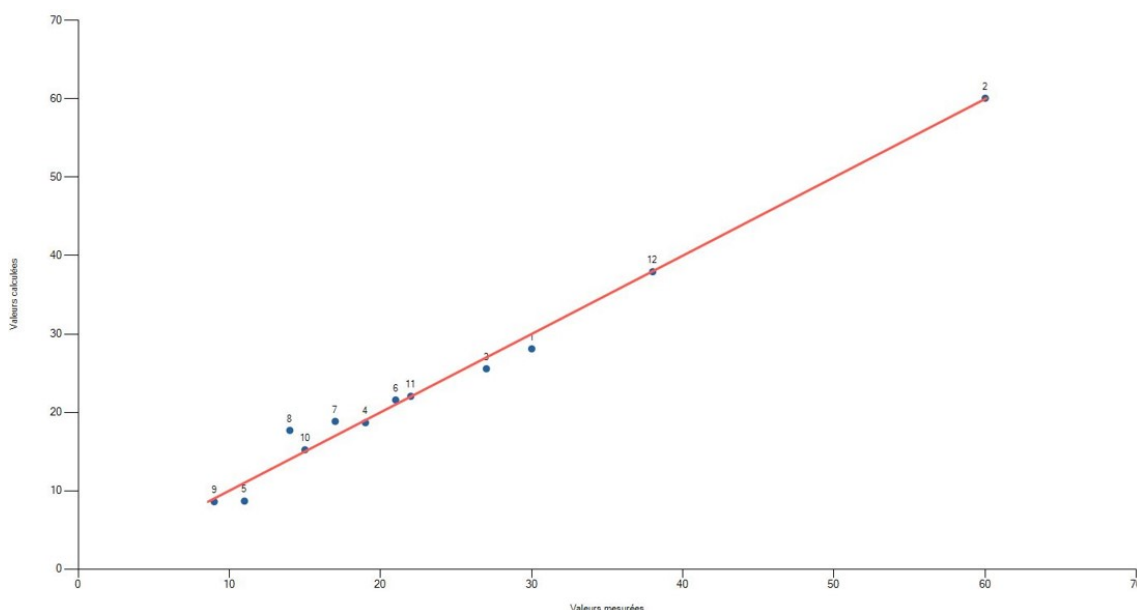


Figure 2.27 – Valeurs d'angles de friction mesurées vs valeurs calculées via le modèle multi-linéaire mis au point

2.5 Conclusion

Ce chapitre nous a tout d'abord permis de présenter un tour d'horizon des principales techniques utilisées dans la littérature pour la description d'images. Elles peuvent être regroupées en 4 catégories :

- les approches géométriques ou de forme, qui permettent de caractériser le contenu morphologique de l'image, c'est-à-dire son organisation.
- les approches statistiques ou de distribution, qui permettent de caractériser la distribution des valeurs des pixels de l'image.
- les approches par filtrage ou par transformation, qui permettent de décrire une image au travers de transformations linéaires ou non, dans les domaines spatial, fréquentiel ou spatio-fréquentiel.
- les approches par modèles, qui permettent de caractériser les images au travers de modèle particulier tels que les fractales ou les champs aléatoires de Markov.

A la suite de cet état des lieux, nous avons introduit l'utilisation de la méthode de Décomposition Modale Discrète (DMD) comme un nouvel outils de description d'image. Cette proposition constitue un apport original de ce travail de thèse.

Après en avoir présenté les principes et les applications existantes, nous avons proposé deux approches pour extraire des descripteurs via cette méthode :

- l'approche coefficients modaux,
- l'approche paramètres statistiques modaux (ou de spectre modal).

Nous avons exposé les résultats d'une étude montrant que la DMD permet d'extraire une information pertinente dans le cadre d'une application de classification.

Le choix des descripteurs à utiliser pour une application donnée est une problématique non aisée à résoudre. En effet, certains attributs sont plus adaptés que d'autres pour décrire certains types d'images ou pour fonctionner avec certaines contraintes liées à l'application, et inversement. C'est ce que montre notamment Mäenpää dans une étude comparative des performances de classification pour différents descripteurs utilisés et différentes conditions d'application, notamment liées aux conditions d'éclairages et aux types de textures d'images à classer. [Mäenpää2004].

De plus, ce choix des descripteurs peut influencer la qualité de la classification. En effet, en plus du fait que tous les descripteurs n'apportent pas forcément d'information en fonction de la tâche de classification visée, leur nombre peut également influencer les résultats. Si ils sont trop peu nombreux, la description des images peut alors être incomplète, engendrant alors de mauvaises performances de classification. Si ils sont trop nombreux, les descripteurs peuvent présenter des corrélations illustrant une redondance de l'information, qui peut également entraîner une dégradation des résultats.

Dans le cadre du développement de systèmes d'inspection automatisée génériques, il est important de pouvoir décrire différents types de produits et différentes classes de défauts, pas forcément connues au préalable, avec le même système. Il est donc essentiel de ne pas limiter la description à une seule catégorie de descripteurs, afin de maximiser les possibilités en terme d'extraction d'information. Cependant, il est également important de réduire le temps de calcul en ne gardant que le minimum de paramètres possible pour une tâche de classification donnée au final, afin de pouvoir s'adapter aux besoins de productivité de la phase de contrôle en ligne industriel.

Afin de répondre conjointement à ces besoins, il est nécessaire de passer par une phase de sélection des descripteurs les plus pertinents parmi tous ceux extraits en phase de description d'images. Il s'agit de la deuxième étape de la réduction d'information avec la classification. Le prochain chapitre traite des moyens qu'il est possible de mettre en œuvre pour réaliser ce tri.

3

Réduction de l'information des images : Tri des descripteurs

Sommaire

3.1	Introduction	79
3.2	Sélection de descripteurs	80
3.2.1	Intérêt dans le contexte de nos travaux	80
3.2.2	Principes généraux	80
3.2.3	Procédure de génération	82
3.2.3.1	Génération complète ou exhaustive	82
3.2.3.2	Génération séquentielle	84
3.2.3.3	Génération aléatoire	85
3.2.3.4	Aide au choix d'un algorithme	87
3.2.4	Fonction d'évaluation	89
3.2.4.1	Mesures de distances	90
3.2.4.2	Mesures d'information	94
3.2.4.3	Mesures de dépendances	95
3.2.4.4	Mesures de consistance	96
3.2.4.5	Mesures du taux d'erreur de classification	96
3.2.5	Critère d'arrêt	97
3.2.6	Validation	98
3.3	Transformation de l'espace des descripteurs	99
3.3.1	Analyse en Composantes Principales (ACP)	99
3.3.2	Analyse en Composantes Indépendantes (ACI)	101
3.3.3	Analyse Linéaire Discriminante (LDA)	102
3.3.4	Algorithme t-SNE	103
3.3.5	Synthèse et mise en perspective dans la cadre du contrôle visuel automatisé en industrie	103
3.4	Critère de séparation basé sur le test du T^2	104
3.4.1	Test du T^2	104
3.4.1.1	Principes du test statistique	105
3.4.1.2	Statistique de test du T^2	105
3.4.2	Application aux cartes de contrôle	106

3.4.3 Critère d'évaluation de distance par le T^2	107
3.5 Conclusion	110

3.1 Introduction

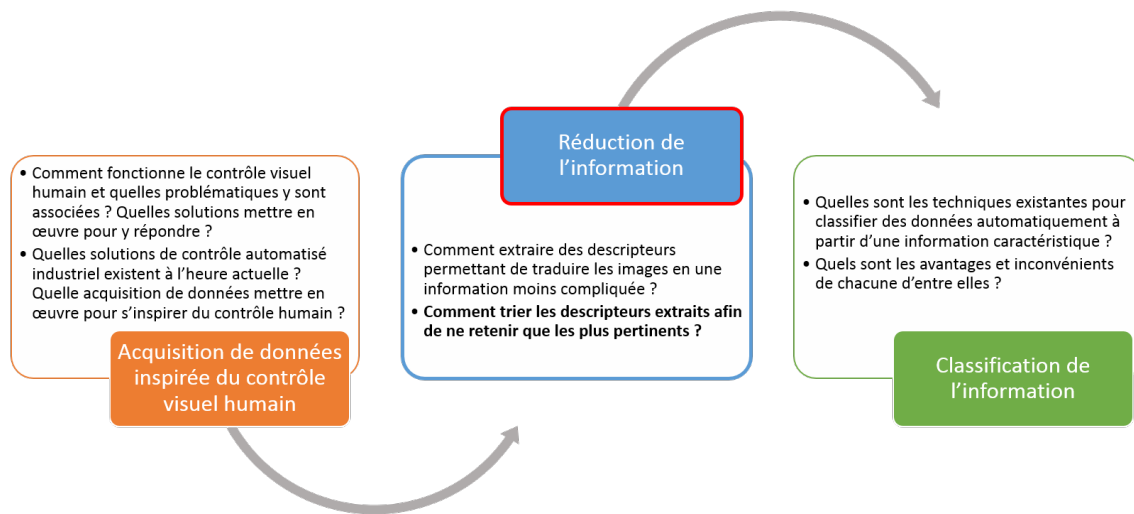


Figure 3.1 – Problématiques liées à la réduction de l'information par la sélection de descripteurs

La description des images décrite dans le chapitre précédent mène souvent à l'extraction de nombreux descripteurs pour tenter de représenter les images le mieux possible. Néanmoins, certains paramètres obtenus ne sont pas pertinents pour traiter la problématique de classification visée, c'est-à-dire n'améliorent pas les résultats, ou se révèlent redondants, donc n'apportent rien de nouveau pour la classification. Ces deux types de caractéristiques ne sont donc pas utiles et il est donc intéressant de pouvoir les éliminer de l'ensemble des descripteurs. Cette suppression a deux effets majeurs.

Le premier est que la performance de classification s'en voit améliorée car seuls les descripteurs apportant une information pertinente sont traités. De plus, le fait d'entraîner un modèle sur un nombre réduit de paramètres lui permet globalement de converger vers des concepts plus généraux, donc de mieux généraliser par la suite sur de nouveaux éléments.

Le second est que la réduction de la taille de l'ensemble des caractéristiques permet de réduire considérablement les temps de calculs pour la classification. Dans le contexte industriel qui nous intéresse, ce second point est très important.

Pour répondre à ce besoin de réduction du nombre de descripteurs, deux approches sont généralement employées : la sélection des descripteurs les plus pertinents et la transformation de l'espace des descripteurs en un espace de plus petite dimension.

Dans ce chapitre, nous proposons de présenter dans un premier temps les principes généraux des techniques de sélection de descripteurs et de réduction de dimensionnalité par transformation de l'espace, puis d'exposer le fonctionnement d'une technique que nous avons développée dans le cadre de ces travaux de thèse.

De manière générale, ce chapitre cherche donc à répondre à la question : "Comment trier les descripteurs extraits afin de ne retenir que les plus pertinents ?".

3.2 Sélection de descripteurs

Les techniques utilisées pour sélectionner les descripteurs les plus pertinents à partir de l'ensemble des descripteurs extraits sont regroupées sous l'appellation de méthodes de *sélection de descripteurs*, ou *sélection de paramètres*, ou encore *sélection d'attributs*. Après une justification de l'intérêt de l'utilisation de ces méthodes dans le cadre de nos travaux, nous proposons de présenter les principes généraux de fonctionnement et les différentes familles de techniques existantes. Nous reprenons notamment pour cette étude des éléments présentés dans des articles d'état de l'art rédigés par Dash et Liu [Dash1997], Doak [Doak1992], Kudo [Kudo2000] et dans une partie de la thèse de Porebski [Porebski2009], qui exposent de manière claire et complète les tenants et aboutissant de ces techniques, ainsi que les différentes grandes catégories d'algorithmes existants.

3.2.1 Intérêt dans le contexte de nos travaux

Dans le cadre de ces travaux de thèse, nous proposons d'exploiter une information multi-éclairages afin de développer un système d'inspection d'aspect automatisé dans un contexte de contrôle en-ligne industriel. Ce système se veut flexible quant aux anomalies à détecter et aux géométries de pièces à inspecter.

L'acquisition multi-éclairages permet d'obtenir une grande quantité d'information qu'il est nécessaire de réduire pour pouvoir l'exploiter efficacement dans un contexte de contrôle en-ligne. Le chapitre précédent a donc porté sur une première phase de réduction de cette information par la description des images. Or, afin de décrire les pièces le plus pleinement possible pour détecter le plus d'anomalies possible sur une variété de types d'objets, nous préconisons d'extraire différents types de descripteurs. Cela a pour conséquence l'obtention d'un nombre de descripteurs encore important à la sortie de cette étape. De plus, tous les descripteurs extraits ne sont pas forcément pertinents pour décrire chaque type de pièces avec les anomalies présentes à sa surface. Il apparaît alors comme nécessaire de procéder à une sélection d'un sous-groupe de descripteurs avant d'effectuer la dernière étape de classification. Cette sélection doit permettre d'obtenir un groupe de descripteurs de taille assez faible pour réaliser la classification dans un temps concordant avec les contraintes de contrôle en-ligne, et contenant une information pertinente pour représenter les anomalies rencontrées.

Partant de cette problématique, nous proposons donc dans la suite de cette partie un tour d'horizon des différentes familles de méthodes communément utilisées pour réaliser ce type de sélection.

3.2.2 Principes généraux

Si plusieurs définitions de la sélection de paramètres existent, Dash et Liu en propose une considérant les différents aspects habituellement abordés [Dash1997]. Pour eux, la sélection d'attributs tente de retenir un sous-ensemble de descripteurs de taille minimale suivant le double critère suivant :

1. La performance de classification ne diminue pas significativement ;
2. La distribution des classes résultantes, donnée uniquement par les paramètres sélectionnés, est aussi proche que possible que la distribution originale, donnée par tous les paramètres.

De manière idéale, les méthodes cherchent le meilleur sous-ensemble d'attributs selon un critère donné, parmi les $2^N - 1$ espaces candidats, N étant le nombre de paramètres de l'ensemble de départ. Néanmoins, dans ce cas la recherche exhaustive de la seule meilleure combinaison est très coûteuse en temps de calcul, et les méthodes utilisées en pratique sont souvent des heuristiques

recherchant un compromis entre la réduction de dimensionnalité et le temps de calcul. Elles requièrent donc un *critère d'arrêt* pour stopper la recherche lorsque ce compromis est atteint.

Inspirés par ces approches, Dash et Lui proposent 4 étapes de bases composant un système de sélection classique, représentées en Figure 3.2 [Dash1997] :

- Une **procédure de génération** qui génère un sous-ensemble. La génération peut commencer à partir de trois situations : pas de paramètres, tous les paramètres, un sous-ensemble aléatoire de paramètres. Dans les deux premiers cas, des paramètres sont ajoutés ou retirés de manière itérative. Dans le dernier cas, des paramètres peuvent être ajoutés ou retirés également, ou un autre sous-ensemble aléatoire peut être créé à chaque itération.
- Une **fonction d'évaluation** qui évalue la performance du sous-ensemble. Elle permet d'évaluer la pertinence du sous-ensemble considéré, qui est ensuite comparée à celui du sous-ensemble précédent. Dans le cas où l'indicateur de pertinence est meilleur, ce sous-ensemble remplace le précédent.
- Un **critère d'arrêt** qui permet de stopper l'algorithme. Plusieurs types de critères existent dont les plus communs sont : l'arrêt a lieu lorsque (i) un certain nombre de caractéristiques sont sélectionnées, (ii) un certain nombre d'itérations ont été effectuées, (iii) l'addition ou la soustraction de nouveaux paramètres n'améliore pas la performance du sous-ensemble, (iv) un sous-ensemble optimal au regard d'un certain critère a été atteint.
- Une **procédure de validation** qui vérifie que le sous-ensemble est bien valide. Cette dernière étape ne fait pas réellement partie du processus de sélection en lui-même, mais elle est essentielle pour vérifier que la sélection est bien optimale. Cette validation peut être réalisée par exemple en comparant les résultats avec d'autres résultats précédemment établis avec d'autres algorithmes de sélection.

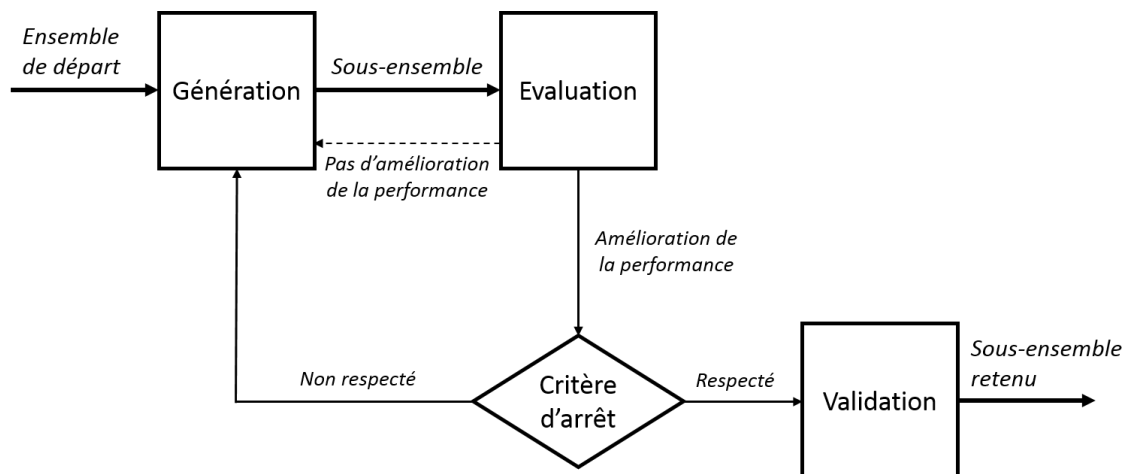


Figure 3.2 – Processus global pour la sélection de paramètres

Après avoir exposé les fondements du processus de sélection de paramètres, nous proposons de détailler un peu plus les différentes étapes du processus, en décrivant pour chacune d'elles les méthodes les plus couramment utilisées. Il ne s'agit pas de fournir un état des lieux exhaustif de toutes les techniques existantes, mais plutôt de donner un aperçu au lecteur des principes de fonctionnement de chaque étape au travers d'algorithmes caractéristiques.

3.2.3 Procédure de génération

Comme indiqué précédemment, il existe $2^D - 1$ combinaisons possibles pour créer des sous-espaces à partir d'un ensemble de D paramètres. Même pour des ensembles de taille moyenne, cela représente un très grand nombre, et il est donc nécessaire d'utiliser des procédures de génération permettant de balayer les combinaisons le plus efficacement possible.

De manière générale, il est possible de regrouper ces approches en 3 catégories : complète, séquentielle et aléatoire.

3.2.3.1 Génération complète ou exhaustive

Les procédures de génération complète, ou exhaustives, recherchent le sous-ensemble optimal, suivant une fonction d'évaluation donnée. Le qualificatif *complet* n'a pas ici pour signification *exhaustif*. En effet, tous les $2^D - 1$ sous-ensembles ne sont pas testés, mais la génération du sous-espace optimal est garantie grâce à l'utilisation de procédures de *backtracking*. Le backtracking regroupe en effet un ensemble de techniques permettant d'abandonner un sous-ensemble candidat à partir du moment où il a été déterminé qu'il ne pouvait pas mener à une solution valide. Parmi les méthodes les plus communes se trouvent :

- l'algorithme de *séparation et évaluation* ou *branch and bound* en anglais,
- l'algorithme du *meilleur en premier*, ou du *best first* en anglais,
- l'algorithme de *recherche par faisceau* ou *beam search* en anglais.

De manière générale, ces algorithmes permettent d'explorer un arbre ou graphe de solutions en ne développant que les branches menant à des solutions prometteuses selon une règle spécifique d'évaluation. Afin d'aborder simplement les principes des méthodes de recherche exhaustive, nous proposons d'expliquer brièvement le fonctionnement global de la famille d'algorithmes *branch and bound* (BB), initialement introduit par Narendra & al [Narendra1977].

Dans cet algorithme, la partie *séparation* ou *branch* a pour objectif de diviser le problème global en différents sous-problèmes permettant de déboucher sur des solutions réalisables. Les ensembles de solutions ainsi obtenus sont souvent représentés graphiquement sous forme d'arbre, nommé *arbre de décision* ou *de solutions*.

La partie *évaluation* ou *bound* consiste en l'exploration de l'arbre de décision en évaluant si une branche mène à une solution valable ou non par rapport à une fonction d'évaluation F donnée. Si une branche est identifiée comme ne pouvant pas mener à une solution acceptable, il n'est alors pas nécessaire de la calculer entièrement. Pour que l'algorithme fonctionne, il est nécessaire que la fonction d'évaluation soit monotone et que la dimension du sous-espace d'attribut final soit connu *a priori*.

Nous proposons un exemple particulier, inspiré d'une illustration issue de la thèse de Porebski [Porebski2009], pour expliciter plus simplement ce mécanisme.

Considérons le problème de sélection de 3 descripteurs dans un espace de 6 descripteurs notés $D_i (i \in 1, 2, 3, 4, 5, 6)$, dont l'arbre de décision issu de la phase de séparation est décrit partiellement en Figure 3.3. Chaque branche représente un sous-espace de solutions dans lequel un descripteur est éliminé à chaque étape, jusqu'à obtenir un nœud contenant le nombre de descripteurs attendus, que nous nommons nœud final.

Nous considérons ici que la fonction d'évaluation F est monotone croissante, c'est-à-dire que pour un nœud N_1 contenant d descripteurs, et un nœud N_2 contenant $d - j$ descripteurs, $F(N_1) > F(N_2)$.

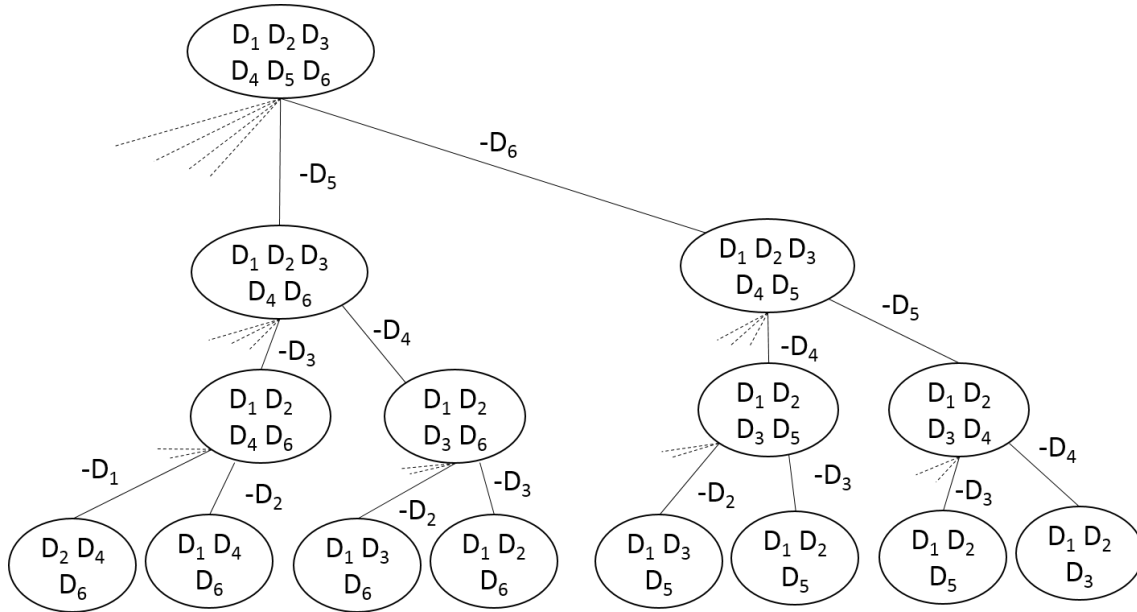


Figure 3.3 – Arbre de décision partiel d'un problème de sélection *branch and bound* pour le choix de 3 descripteurs parmi 6

La procédure de sélection se déroule alors selon les deux grandes étapes suivantes :

1. La procédure de sélection commence par le calcul de la valeur de la fonction d'évaluation pour un premier chemin menant à un nœud final quelconque. Cette valeur, nommée F_{max} , correspond alors au meilleur nœud final évalué.
2. Le reste de l'arbre est évalué par rapport à cette valeur initiale de F_{max} . La valeur de F est donc calculée pour chaque chemin menant à un nœud final. Néanmoins, tous les chemins ne sont pas forcément suivis jusqu'au bout, grâce à un procédé de *backtracking*, permis par la monotonie de la fonction d'évaluation. En effet, si la valeur de la fonction d'évaluation pour un nœud parent est inférieure à F_{max} , il n'est alors pas nécessaire d'évaluer les autres solutions de cette branche, car elles donneront des performances forcément plus faibles également. A chaque fois qu'un nœud final est évalué comme meilleur que le celui précédemment retenu, F_{max} est mis à jour.
3. Une fois tous les nœuds finaux valables testés, le dernier retenu contient donc la meilleure combinaison de descripteurs.

Si cet algorithme permet de trouver la meilleure solution parmi toutes celles existantes à partir de l'espace de départ, il affiche cependant certains inconvénients :

- Ce type d'algorithme est très demandeur en temps de calcul, qui augmente d'ailleurs très rapidement avec le nombre de descripteurs de l'espace initial. Si il existe bien certaines techniques et stratégies pour réduire ce temps, un étude menée par Kudo recommande de ne pas les utiliser pour les espaces de plus de 50 descripteurs [Kudo2000].
- Il s'avère généralement difficile, voire impossible, de connaître par avance la dimension du sous-espace de descripteurs optimal. La solution la plus simple est de fixer arbitrairement un nombre de dimensions pour le sous-espace final, au risque de se couper de certaines solutions performantes. Une autre possibilité est de chercher un sous-espace de taille minimale, mais garantissant une capacité de discrimination ne descendant pas en dessous d'un seuil fixé par l'utilisateur.
- Le besoin d'une fonction d'évaluation monotone est également une difficulté majeure pour l'application de cette méthode. La première raison est que cet aspect monotone a tendance

à orienter la sélection vers des sous-ensembles de descripteurs d'assez grandes dimensions [Doak1992]. La seconde raison est que certains problèmes présentent des fonctions de décision non monotones. L'algorithme ne peut donc pas être appliqué dans leur cas.

De manière générale, les algorithmes de recherche exhaustive présentent une grande précision et permettent de trouver généralement la solution optimale. Néanmoins, ils deviennent rapidement très complexes et ils sont donc difficiles à mettre en œuvre pour des espaces de descripteurs initialement de grandes dimensions (>50).

Cependant, un bon compromis peut être d'utiliser tout d'abord un autre type d'algorithme (séquentiel ou aléatoire), moins complexe, pour réduire le nombre de descripteurs, puis d'utiliser un algorithme de recherche exhaustive sur l'ensemble réduit [Doak1992].

Afin de réduire la complexité de la génération, et ainsi pouvoir traiter des ensembles de descripteurs plus grands, il existe des méthodes sous-optimales, c'est-à-dire ne menant pas forcément à la solution optimale car n'explorant pas tout l'espace des possibilités. Il s'agit des catégories de générations séquentielles et aléatoires.

3.2.3.2 Génération séquentielle

Les algorithmes de génération séquentielle ont pour objectif d'explorer seulement une petite partie des solutions possibles. Leur principe global est d'ajout ou/et de suppression de descripteurs suivant un processus itératif.

Il existe plusieurs méthodes pour ce type de génération, parmi lesquelles il est possible de citer :

- Les méthodes de type *Forward* (FSS) qui consistent en l'ajout de descripteurs à un ensemble vide ;
- Les méthodes de type *Backward* (BSS) qui consistent en la suppression de descripteurs à partir de l'ensemble complet ;
- Les méthodes de type (p,q) (PQSS) qui consistent en l'ajout de la meilleure combinaison de p descripteurs suivi de la suppression de la pire combinaison de q descripteurs ;
- Les méthodes de type flottantes, soit *Forward floating* (SFFS), soit *Backward floating* (SBFS), extensions des méthodes (p,q) , qui consistent en l'application de plusieurs étapes *Backward* après une étape *Forward* (ou inversement) tant que le sous-espace obtenu améliore la valeur de la fonction d'évaluation ;
- Les méthodes de type *Bi-Directional* (BDS) qui consistent en l'utilisation de deux algorithmes de types FSS et BSS en parallèle pour trouver un sous-espace intermédiaire.

Nous proposons de présenter un peu plus en détails les approches FSS et BSS qui sont les plus communément référencées et qui servent de base aux autres approches.

L'approche de sélection séquentielle *Forward* fonctionne selon les étapes suivantes :

1. Choix d'une fonction d'évaluation F pour sélectionner des descripteurs parmi un ensemble initial de dimension D . Imaginons par exemple que cette fonction soit à maximiser pour obtenir la meilleure capacité de discrimination.
2. Initialisation du sous-espace à évaluer à un ensemble vide. Soit d la dimension de ce sous-espace, d vaut donc initialement 0.
3. La valeur de d est incrémentée de 1 jusqu'à $d = D$ et à chaque itération, un descripteur est choisi parmi ceux non encore sélectionnés pour être ajouté au sous-ensemble existant. Le descripteur choisi à chaque fois est celui qui, en combinaison avec ceux déjà présent dans le sous-ensemble, permet de maximiser la fonction d'évaluation pour le nouveau sous-ensemble formé.

4. L'ajout de descripteur s'arrête lorsqu'un critère d'arrêt est atteint. Ce critère d'arrêt peut par exemple porter sur un nombre maximum de descripteurs à atteindre ou encore sur une augmentation minimale de la valeur de la fonction d'évaluation d'une itération à une autre. Les différents types de critères d'arrêts sont développés un peu plus loin dans ce chapitre.

L'approche *Backward* est similaire à celle décrite ci-dessus pour le *Forward* à l'exception que le sous espace de départ est initialisé comme contenant tous les descripteurs, et qu'un descripteur est supprimé à chaque itération.

Ces méthodes sont beaucoup moins complexes et coûteuses en temps de calcul que les approches de génération exhaustives car elles n'explorent pas toutes les solutions possibles. Cependant, dans le cas du processus *Forward*, l'ajout de nouveaux descripteurs peut rendre obsolète certains descripteurs ajoutés précédemment. De même pour le processus *Backward*, le retrait de nouveaux descripteurs peut rendre utile un descripteur retiré précédemment. Or, l'impossibilité de retirer ou d'ajouter à nouveau un descripteur après qu'il ait été ajouté ou retiré du sous-espace étudié, peut avoir pour conséquence de garder un paramètre obsolète ou de se priver d'un paramètre apportant au final de l'information. Cette incapacité de retour en arrière a pour conséquence que l'algorithme ne mène au final pas forcément à la meilleure solution.

Pour tenter de contrer cet effet et ainsi d'améliorer les performances de ce type d'algorithmes, d'autres approches citées précédemment cherchent à coupler ces deux approches élémentaires afin d'explorer plus de possibilités.

Les approches de type (p,q) ont par exemple pour principe de réaliser tout d'abord p fois une procédure *Forward*, puis d'enchaîner avec q fois une procédure *Backward*. Les valeurs de p et q sont fixées par l'utilisateur.

Les approches flottantes sont des extensions de l'approche (p,q) dans lesquelles une itération *Forward* (ou *Backward*) est réalisée et suivie d'une ou plusieurs itérations *Backward* (ou *Forward*) tant que le sous-espace obtenu améliore la valeur de la fonction d'évaluation.

Ces types de fonctionnements intègrent donc une rétroaction dans le processus, permettant ainsi d'obtenir des solutions plus performantes, mais augmentant d'un autre côté la complexité et le temps de calcul.

Elles s'avèrent néanmoins être un bon compromis et sont considérées comme les méthodes sous-optimales les plus efficaces [Porebski2009].

L'autre famille d'approches sous-optimales est la génération aléatoire, dont les principes sont résumés ci-après.

3.2.3.3 Génération aléatoire

Les approches de génération aléatoire permettent de parcourir en partie au hasard une fraction de l'ensemble des solutions. L'introduction de hasard dans le choix des solutions évaluées permet d'éviter de rester dans des minimas d'optimisation locaux. L'hypothèse à la base de ces approches est qu'en acceptant de quitter périodiquement un chemin d'optimisation menant à de meilleurs résultats pour un état présentant de moins bons résultats, l'algorithme évite de tomber dans un minima local.

Parmi les algorithmes de génération aléatoires référencés, il est notamment possible de citer :

- Les algorithmes *génétiques* (AG), introduits par Holland [Holland1975] et inspirés de la théorie de la sélection naturelle ;
- Les algorithmes de *recuit simulé* ou *simulated annealing* (SA) en anglais, adaptés d'une méthode initialement proposée par Metropolis [Metropolis1953] et inspirés d'un processus

utilisé en métallurgie.

- Les algorithmes de *génération aléatoire plus sélection séquentielle* ou *Random Generation Plus Sequential Selection* (RGSS), qui consiste en la génération aléatoire d'un sous-espace de descripteurs, suivie de l'application d'un FSS et d'un BSS, permettant ainsi d'élargir l'espace de recherche des approches séquentielles [Doak1992].

Nous proposons ci-dessous de détailler un peu plus le fonctionnement des algorithmes génétiques, qui sont globalement les plus utilisés.

Les algorithmes génétiques utilisent une représentation des données sous forme de vecteurs appelés *chromosomes*, composés de *gènes* pouvant prendre les valeurs 0 ou 1. Chaque vecteur contient D gènes, D représentant le nombre de descripteurs total dans l'ensemble étudié. Si le gène numéro i , $i \in (0, 1, 2, \dots, D)$, prend la valeur 1, cela signifie que le descripteur D_i est retenu dans la combinaison de descripteurs codée par ce chromosome. Si le gène est à 0, cela signifie que le descripteur n'est pas retenu dans la combinaison.

Le fonctionnement de ce type d'algorithme est illustré en Figure 3.4 et décrit ci-après.

L'initialisation de l'algorithme se fait en générant N chromosomes aléatoirement à partir de l'ensemble des descripteurs de départ. Le processus qui suit est itératif et permet de générer une nouvelle population à chaque itération, plus "adaptée" au sens de la fonction d'évaluation. L'évolution de cette population est réalisée grâce à des traitements dits génétiques, correspondant à 3 opérateurs :

1. La *sélection* : elle consiste tout d'abord à classer les chromosomes selon leur performance de discrimination par rapport à la fonction d'évaluation. Puis, les $M = N \times \alpha$ premiers meilleurs chromosomes sont gardés pour la suite du traitement. α est un pourcentage fixé par l'utilisateur.
2. Le(s) *croisement(s)* : cette étape consiste à croiser un certain nombre de couples de chromosomes "parents" pour créer des couples de chromosomes "enfants". Chaque couple parent donne naissance à un couple enfant, selon le mécanisme illustré en Figure 3.4. Seul un certain pourcentage β de chromosomes est croisé, paramètre fixé par l'utilisateur.
3. La(es) *mutation(s)* : cette dernière phase consiste à modifier la valeur de certains gènes de manière aléatoire, comme illustré en Figure 3.4. Cela revient à ajouter ou retirer un descripteur d'une combinaison considérée, permettant ainsi d'introduire de la diversité dans l'ensemble des solutions. Seul un pourcentage ν de gène est modifié, dont la valeur est fixée par l'utilisateur.

Ces 3 étapes sont répétées jusqu'à ce qu'un critère d'arrêt soit atteint comme pour les algorithmes séquentiels.

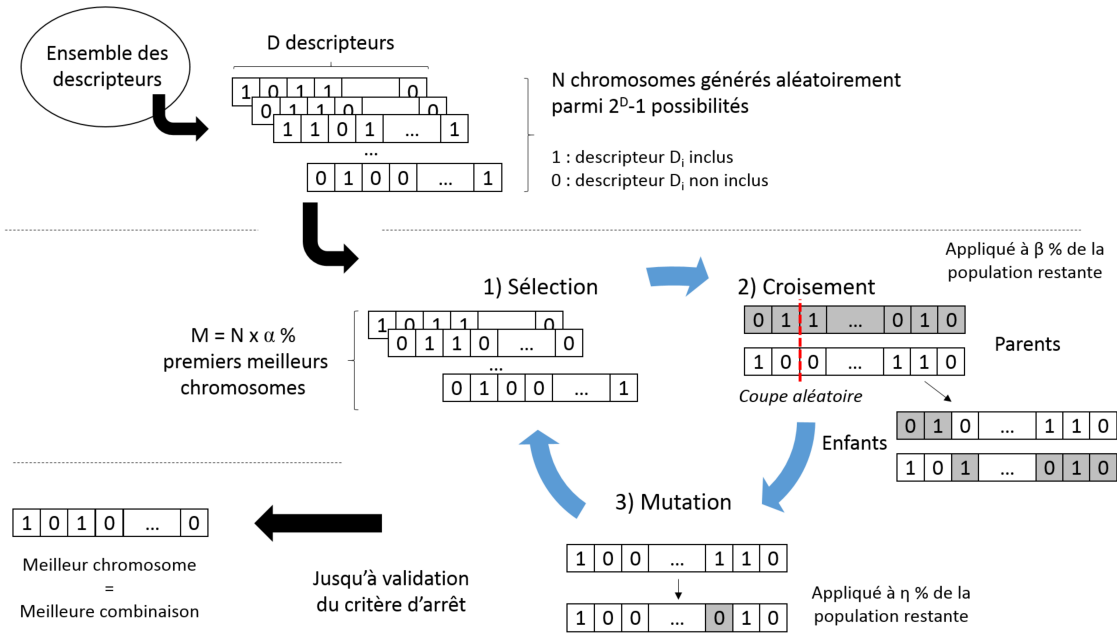


Figure 3.4 – Illustration du fonctionnement d'un algorithme de sélection génétique

Comme pour les approches séquentielles, ce type d'algorithme a pour avantage d'être moins complexe et moins consommateur en temps de calcul que les approches exhaustives. De plus, elles ne requièrent pas de fonction d'évaluation monotone.

Après avoir décrit les différents types d'approches pour la génération de sous-espaces d'attributs, nous proposons maintenant de résumer les forces et limites de ces différentes approches, ainsi que certains conseils donnés dans la littérature pour effectuer un choix entre elles en fonction du problème considéré.

3.2.3.4 Aide au choix d'un algorithme

Nous proposons pour terminer d'exposer quelques conseils pour effectuer un choix entre les différents types d'algorithmes cités précédemment. Pour cela, nous reprenons des résultats et remarques principalement issus des études de Kudo [Kudo2000], Doak [Doak1992] et Porebski [Porebski2009].

La pertinence de l'utilisation des différents algorithmes est étudiée selon l'objectif visé, selon la taille de l'ensemble initial des descripteurs et le type de fonction d'évaluation utilisés. Nous proposons également de rappeler en parallèle certaines considérations portant sur la complexité et la précision des différents groupes de méthodes.

Les 3 types d'objectifs considérés sont :

- Trouver le meilleur sous-espace en terme de capacité de discrimination, pour une taille de sous-espace d'arrivée fixée.
- Trouver le plus petit sous-espace possédant une capacité de discrimination fixée.
- Trouver un sous-espace permettant un compromis entre la taille du sous-espace d'arrivée et la capacité de discrimination.

Le tableau 3.1 expose une classification des algorithmes BB (exhaustif), SFFS et SBFS (séquentiels) et AG (aléatoire) suivant les différents critères cités précédemment.

	D < 20		20 ≤ D ≤ 50		50 ≤ D ≤ 100		D > 100
Objectif	M	NM	M	NM	M	NM	
A	BB	*	BB	*	AG SFFS SBFS	*	SFFS AG
B	BB	*	BB AG	*	AG	*	SFFS AG
C	*	AG SFFS SBFS	*	AG SFFS SBFS	*	AG	SFFS AG

Tableau 3.1 – Récapitulatif des algorithmes de génération en fonction de critères situationnels (* : situations rarement rencontrées donc non étudiées pour cette classification - M : Fonction d'évaluation Monotone - NM : Fonction d'évaluation Non Monotone) [Kudo2000] [Porebski2009]

A partir de cette étude comparative issue de la littérature, nous proposons quelques remarques générales sur les caractéristiques des différents groupes de méthodes et leur utilisation.

Tout d'abord, il est possible de remarquer que lorsque la fonction d'évaluation est monotone, la plupart des problèmes rencontrés cherchent à résoudre des objectifs de type A ou B, c'est-à-dire permettant de fixer soit la dimension, soit le seuil de discrimination à atteindre.

Dans le cas d'une fonction d'évaluation non monotone, les problèmes sont en général plus compliqués à résoudre et c'est un objectif de type C qui est souvent visé (compromis entre la dimension et le seuil de discrimination).

Les approches de génération complète sont plutôt adaptées pour des problèmes cherchant à répondre à des objectifs de type A ou B, présentant des fonctions d'évaluation monotones et des ensembles de descripteurs de départ de dimension assez petite (<50). En effet, comme expliqué précédemment, elles permettent de trouver dans tous les cas la solution optimale du problème, mais requiert pour cela un temps de calcul important, augmentant de manière exponentielle avec le nombre de descripteurs à traiter.

Les approches de génération séquentielle sont plutôt adaptées pour des problèmes plus compliqués portant sur des objectifs de type C avec des fonctions d'évaluation non monotones, et pour les 3 types d'objectifs dès que la dimension de l'espace de départ augmente. En effet, n'explorant qu'une partie de l'espace, elles permettent de converger plus rapidement vers une solution pour de larges ensembles de descripteurs. Néanmoins, les processus utilisés ne permettent en général de mener qu'à des solutions sous-optimales.

L'introduction de procédés de *backtracking* avec les méthodes SFFS et SBFS permettent d'améliorer l'optimalité des solutions trouvées, mais augmentent en contrepartie la complexité de l'algorithme pour de grands ensembles de descripteurs. Si ces approches restent plus adaptées que les approches exhaustives quand le nombre de descripteurs augmente, il est donc néanmoins important de garder en tête que leur complexité peut rester importante pour de grands ensembles de données.

Enfin, les approches de génération aléatoire sont globalement adaptées pour être utilisées dans le même type de situations que les approches séquentielles. L'avantage de l'introduction de hasard dans la sélection peut permettre à ces méthodes d'éviter des minimums locaux et donc de converger vers des solutions plus proches de l'optimalité globale. Il a été montré que pour des ensembles de descripteurs de tailles moyennes (20 à 100), les méthodes aléatoires de type AG ont tendance

à converger vers des solutions comparables voire meilleures en performances que des approches séquentielles de type SFFS et SBFS. Néanmoins, lorsque la dimension augmente, les résultats obtenus avec des AG semblent devenir moins bons [Kudo2000].

Par contre, les algorithmes aléatoires présentent un principal avantage par rapport aux autres catégories dans le fait que leur complexité n'augmente pas significativement avec le nombre de descripteurs de l'espace initial. Le temps de calcul de cette famille d'algorithme devient donc bien plus avantageux que ceux des autres familles pour des ensembles de grandes tailles.

Cependant, ce groupe de méthodes présentent généralement un réglage de paramètres (fréquences de croisements et mutations pour les GA par exemple) compliqué à réaliser, et qui peut influencer grandement sur les résultats obtenus [Doak1992].

La sélection d'un sous-ensemble via les méthodes présentées précédemment repose sur l'estimation de la capacité de discrimination de celui-ci grâce à une fonction d'évaluation donnée. Les différents types de fonctions pouvant être utilisées à cet effet sont présentées dans la partie suivante.

3.2.4 Fonction d'évaluation

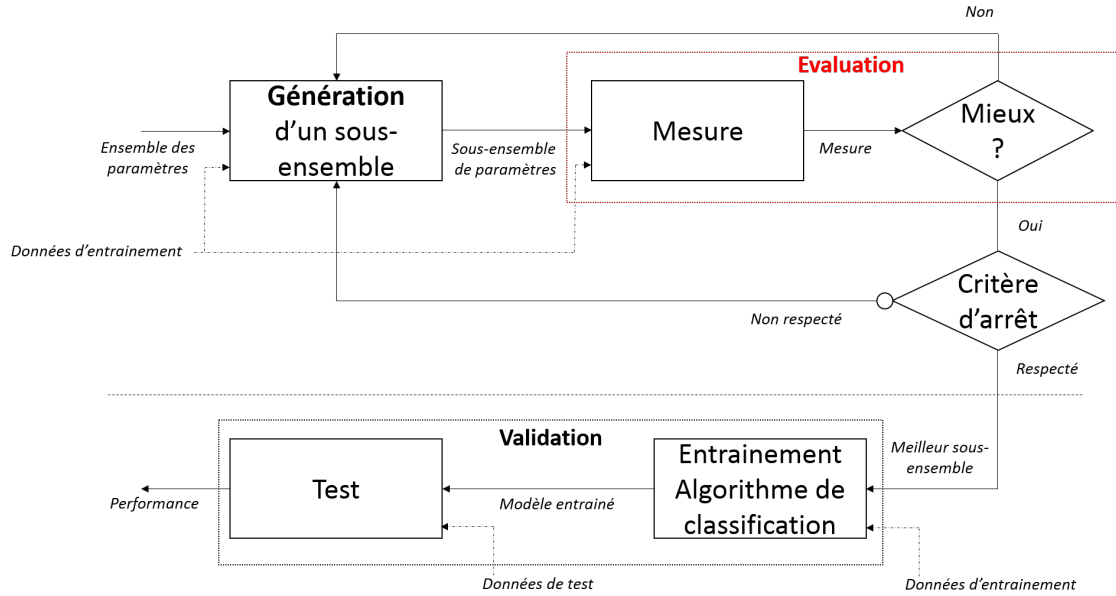
La notion de sous-espace optimal est toujours relative au choix d'une fonction d'évaluation. En effet, un sous-ensemble optimal pour une certaine fonction ne sera probablement pas optimal pour une autre fonction, pour un même ensemble de descripteurs donné. Il est important de bien choisir cette fonction pour que la sélection réponde au problème posé.

Dans le cadre de notre application en classification automatique, une fonction d'évaluation tente généralement de mesurer la capacité d'un sous-ensemble d'attributs à pouvoir séparer des éléments de différentes classes.

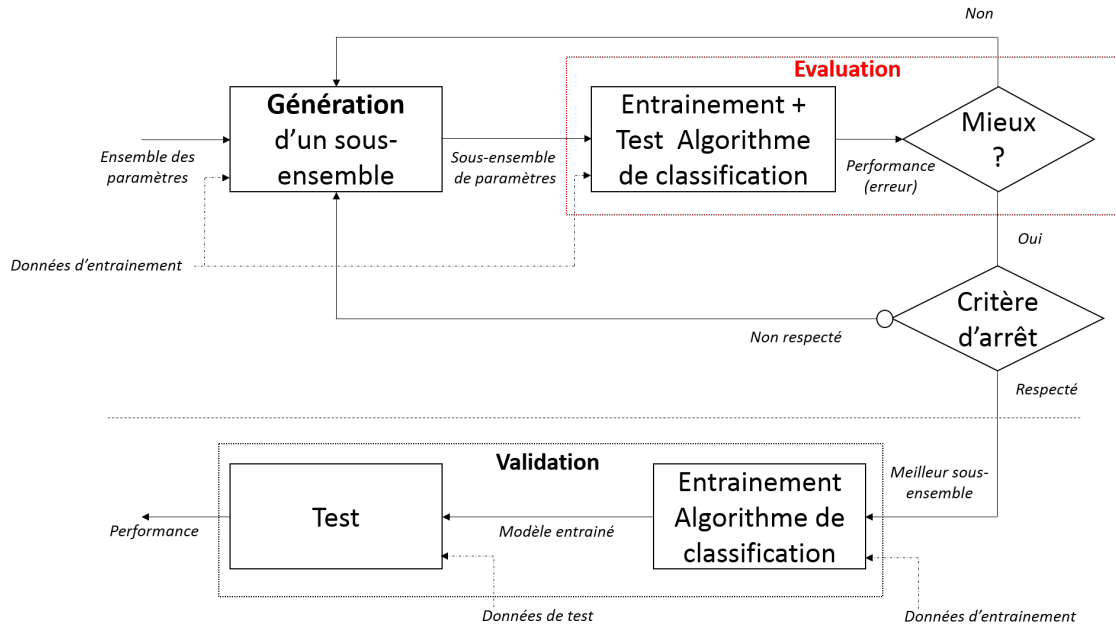
Dash et Lui proposent dans leur étude de regrouper ces fonctions en 5 catégories [Dash1997] :

- Mesures de distances ;
- Mesures d'information ;
- Mesures de dépendances ;
- Mesures de consistance ;
- Mesures du taux d'erreur de classification.

En faisant le parallèle avec une autre classification proposée par John et al., les 4 premières familles font parties des approches dites de type *filter* [John1994]. Cela signifie que l'évaluation des sous-ensembles est réalisée indépendamment de l'algorithme de classification choisi. Contrairement à ces dernières, les mesures du taux d'erreur de classement appartiennent aux approches de type *wrapper*, et sont dépendantes de la méthode de classification utilisée. Ces dernières approches ont tendance à être plus précises car elles fondent la sélection des paramètres directement sur les performances de classification, mais elles ont en contrepartie un coût en temps de calcul beaucoup plus grand [Dash1997]. La Figure 3.5 détaille le processus de sélection pour ces deux types d'approches.



(a) Approche Filter



(b) Approche Wrapper

Figure 3.5 – Illustration du processus de sélection pour les deux types d'approches Filter et Wrapper

3.2.4.1 Mesures de distances

Les mesures de distances, aussi appelées mesures de séparabilité, de divergence ou encore de discrimination, ont pour objectif d'évaluer la capacité d'un sous-ensemble de paramètres à séparer les classes dans l'espace de représentation ainsi généré. Il s'agit alors de rechercher un sous-ensemble de descripteurs tel que :

- la *variance inter-classes* est maximale, c'est-à-dire que les éléments appartenant à des classes différentes forment des nuages de points les plus séparés possibles dans l'espace des descripteurs choisis ;
- la *variance intra-classes* est minimale, c'est-à-dire que les éléments appartenant à des mêmes classes soient les plus proches possibles les uns des autres.

La Figure 3.6 permet de mieux se rendre compte des cas pouvant être rencontrés lors de mesures de distances. Le cas 3.6a illustre une situation où la séparation est bonne, c'est-à-dire que la variance inter-classe est forte et que la variance intra-classe est faible. Le cas 3.6b illustre un cas moins favorable.

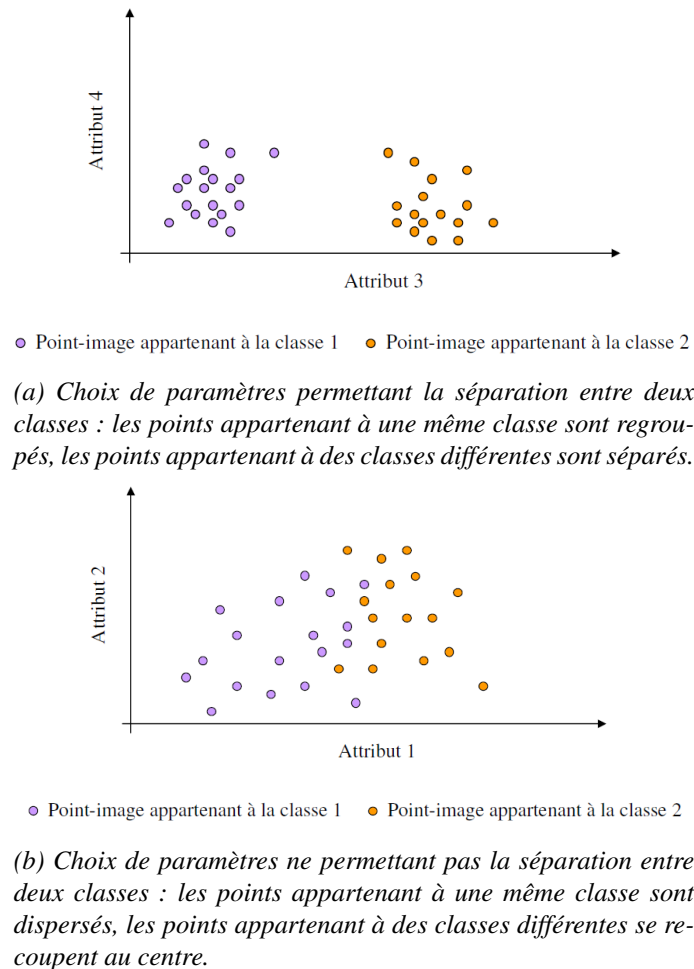


Figure 3.6 – Exemples de sous-ensembles séparateurs ou non du point de vue d'une mesure de distance - issus de [Porebski2009]

Les mesures de distances permettant d'évaluer la discrimination entre les différentes classes peuvent alors être réalisées suivant différentes approches :

- l'évaluation des matrices de covariance intra et inter-classes ;
- l'évaluation de distances mathématiques entre les données.

Critères basés sur les matrices de covariances

L'outil le plus communément utilisé pour évaluer la variance intra et inter-classes est la *matrice de variance-covariance*, ou simplement *matrice de covariance*. Il s'agit d'une matrice carrée de dimensions $N \times N$, avec N le nombre de variables, qui regroupe les variances des différentes variables sur sa diagonale, et les covariances entre les différents couples de variables autour.

La **matrice de covariance intra-classes** W (pour *Within* en anglais) est alors définie selon l'équation suivante :

$$W = \frac{1}{p} \sum_{k=1}^K p_k V_k \quad \text{avec} \quad V_k = \frac{1}{p_k} \sum_{x_i \in C_k} (x_i - g_k)^T (x_i - g_k) \quad (3.1)$$

et avec p le nombre d'individus x total, p_k le nombre d'individus x appartenant à la classe k et g_k le centre de gravité de la classe k défini par l'équation 3.2 .

$$g_k = \frac{1}{p_k} \sum_{x_i \in C_k} x_i \quad (3.2)$$

dans ce cas, chaque x_i se voit attribuer le même poids, mais il est possible de pondérer chaque individu selon son ordre d'importance dans un cas plus général.

La **matrice de covariance inter-classes** B (pour *Between* en anglais) est quant à elle définie par l'équation suivante :

$$B = \frac{1}{p} \sum_{k=1}^K (g_k - g)^T (g_k - g) \quad (3.3)$$

avec g le centre de gravité du nuage de données total défini par l'équation 3.8.

$$g = \frac{1}{p} \sum_i x_i \quad (3.4)$$

Grâce aux équations 3.1 et 3.3, il est possible de définir la matrice de covariance totale V :

$$V = B + W \quad (3.5)$$

Différents critères ont été proposés pour exploiter ces matrices dans le cadre de mesures de distances entre différents groupes de données. Nous reprenons ici quelques exemples des principaux critères utilisés, qui ont notamment été répertoriés par Vandenbroucke [Vandenbroucke2000] :

— le critère **de la trace** :

Il consiste en la comparaison de la matrice de covariance inter-classes avec la matrice de covariance totale, au travers du calcul de la trace du quotient des deux matrices :

$$crit_{Trace} = Tr\left(\frac{B}{T}\right) \quad (3.6)$$

avec l'opérateur trace Tr qui effectue la somme des éléments de la première diagonale d'une matrice carrée.

Plus $crit_{Trace}$ est grand, plus le sous-espace de descripteur considéré est discriminant pour les classes de données étudiées.

— le critère **de Hotelling** :

Ce critère, issu principalement des travaux de Hotelling sur la dispersion des données multivariées [Hotelling1951], est très similaire au précédent. En effet, il consiste à calculer la trace du quotient de la matrice de covariance intra-classes et de la matrice de covariance inter-classe. En d'autre terme, il permet d'étudier le rapport entre la compacité des données dans chaque classe et la séparabilité des différentes classes. Son expression est la suivante :

$$crit_{Hotelling} = Tr\left(\frac{W}{B}\right) \quad (3.7)$$

Ce critère est également à maximiser pour trouver les attributs permettant de discriminer au mieux les classes.

— le critère **de Wilks** :

Ce critère est issu des travaux de Wilks sur l'estimation de paramètres de populations à partir d'échantillons [Wilks1932]. Il consiste à calculer le ratio des déterminants de la matrice de covariance intra-classes et de la matrice de covariance totale :

$$crit_{Wilks} = \frac{\det(W)}{\det(T)} \quad (3.8)$$

Dans ce cas, le sous-ensemble de descripteurs est considéré comme très discriminant si la valeur du critère est faible. On cherche donc à minimiser $crit_{Wilks}$.

— D'autres critères existent mais ne sont pas développés dans ce manuscrit. Nous invitons le lecteur à lire les travaux de thèse de Vandenbroucke qui livre une liste plus exhaustive [Vandenbroucke2000].

Ce genre de critères a par exemple été utilisé par Fiete et Insana pour des applications de traitement d'images et de reconnaissance de paramètres issus de mesures ultrasoniques [Fiete1987] [Insana1986].

Nous avons donc présenté les principes de l'approche de mesures de distances par évaluation de la covariance des données. Ce type de critères permet d'analyser l'ensemble des données de manière globale au travers des matrices calculées.

L'autre groupe d'approches permettant de mesurer des distances entre les classes de données utilisent le calcul de distances dites mathématiques, dont les plus classiques sont présentées dans la suite de cette section.

Distances mathématiques

Les mesures de distances peuvent aussi être réalisées avec le calcul de *distances mathématiques* point à point, sans passer par les matrices de covariance des données. Le Tableau 3.2 référence certaines des distances mathématiques les plus connues.

Les distances les plus considérées dans le cas des applications de sélection de descripteurs sont la distance Euclidienne et la distance de Mahalanobis. Elles sont par exemple utilisées par Bradley [Bradley1998] et Pudil [Pudil1994] comme critère de séparation. La distance de Mahalanobis diffère de la distance Euclidienne dans le fait qu'elle prend en considération la variance des données. Elle accorde en effet un poids moins important aux composantes les plus dispersées.

Distance Euclidienne	$d(x, y) = \sqrt{\sum_{i=1}^n x_i - y_i ^2}$
Distance de Minkowsky	$d(x, y) = (\sum_{i=1}^n x_i - y_i ^p)^{1/p}$
Distance de Manhattan	$d(x, y) = \sum_{i=1}^n x_i - y_i $
Distance de Chebychev	$d(x, y) = \max_{i=1}^n x_i - y_i $
Distance de Camberra	$d(x, y) = \sum_{i=1}^n \frac{ x_i - y_i }{ x_i + y_i }$
Distance de Mahalanobis	$d(x, y) = \sqrt{(x - y)^T \Sigma^{-1} (x - y)}$

Tableau 3.2 – Distances usuelles entre deux points $x = (x_1, x_2, \dots, x_n)$ et $y = (y_1, y_2, \dots, y_n)$ (Σ correspond à la matrice de covariance des données)

Cependant, ce type d'approche est moins référencée que celle utilisant les matrices de covariances, et nous ne la développons donc pas plus dans ce manuscrit.

Après s'être intéressé aux mesures de distances, nous proposons de développer les principes des approches fondées sur les mesures d'information pour évaluer la performance d'un sous-ensemble de descripteurs.

3.2.4.2 Mesures d'information

Les mesures d'information appliquées à l'évaluation de la pertinence d'un sous-ensemble de descripteurs font appel à des principes de la théorie de l'information, qui trouve son origine principale dans les travaux de Claude Shannon et Warren Weaver publiés pour la première fois en 1948 [Shannon2001].

Cette théorie vise de manière générale à quantifier et qualifier le contenu en information présent dans un ensemble de données. Pour mesurer la quantité d'information moyenne d'un ensemble de d'évènements et de caractériser son incertitude, Shannon introduit la notion d'entropie, notée H , dont la définition générale est présentée par l'Equation 3.9. Il existe plusieurs définitions pour le calcul de l'entropie, mais celle proposée par Shannon est la plus commune.

$$H(I) = - \sum_{i \in I} p_i \times \log_2(p_i) \quad (3.9)$$

avec p_i la probabilité d'apparition de l'évènement i appartenant à un ensemble d'évènements I .

La valeur de H est maximale quand les probabilités p_i sont proches les unes des autres, c'est-à-dire que les évènements sont proches de l'équiprobabilité.

Un signal ou un ensemble de données contenant de l'information est en général diversifié, peu prédictible et peu redondant. Son entropie est donc globalement faible et la mesure de cette entropie peut ainsi permettre d'évaluer son contenu informatif.

Dans le cas de l'application de ces principes à la sélection de paramètres pour un problème de classification, le contenu informationnel est évalué à partir des probabilités *a priori* $P(C_k)$, avec $k \in K$ classes, c'est-à-dire les probabilités d'appartenir à la classe C_k . L'Equation 3.9 pour l'entropie devient donc dans ce contexte :

$$H(C_k) = - \sum_{k \in K} P(C_k) \times \log_2(P(C_k)) \quad (3.10)$$

A partir de cette équation, il est possible de définir l'entropie $H(C_k|\vec{D})$, c'est-à-dire l'entropie de C_k étant donné un vecteur de descripteurs $\vec{D} = (D_1, D_2, \dots, D_N)$:

$$H(C_k|\vec{D}) = - \sum_{i \in N} P(D_i) \sum_{k \in K} P(C_k|D_i) \times \log_2(P(C_k|D_i)) \quad (3.11)$$

avec $P(D_i)$ les probabilités *a priori* pour toutes les valeurs de $\vec{D}$ et $P(C_k|D_i)$ les probabilités *a posteriori* d'appartenir à une classe C_k sachant les valeurs de D_i .

La quantité par laquelle l'entropie de C_k diminue représente l'apport d'information apporté par D_i , appelé *gain d'information* et définit par la formule :

$$IG(C_k|D_i) = H(C_k) - H(C_k|\vec{D}) \quad (3.12)$$

Dans le cadre de la sélection de descripteurs par mesure d'information, un sous-ensemble $\vec{D}_1$ de descripteur est préféré à un autre $\vec{D}_2$ si le gain d'information apporté par $\vec{D}_1$ est plus grand que celui apporté par $\vec{D}_2$.

Après s'être intéressé aux mesures d'information, nous présentons dans la prochaine section les principes des approches fondées sur les mesures de dépendances pour évaluer la performance d'un sous-ensemble de descripteurs.

3.2.4.3 Mesures de dépendances

Les mesures de dépendances font référence à des mesures de corrélation entre variables. Ce type de mesure permet de caractériser :

- la corrélation entre un attribut et la répartition des classes. Dans ce cas, si la corrélation entre un descripteur D_1 avec la répartition des classes C est plus forte que celle d'un descripteur D_2 , le premier est préféré au second. En effet, si l'évolution des valeurs d'un descripteur est proche de celle de la répartition des classes pour un lot de données d'apprentissage, alors on considère qu'il apporte une information pertinente.
- la corrélation des descripteurs entre eux. Dans ce cas, la mesure permet de trouver des redondances dans les descripteurs afin de les supprimer.

Ces mesures sont en réalité des cas particuliers de mesures de distances et d'information, mais elles sont tout de même séparées dans une nouvelle catégorie car elles représentent un autre point de vue conceptuel [Dash1997].

Le coefficient de corrélation, dont la définition est présentée en Figure 3.13, est communément utilisé comme indicateur. Plus il est proche de 0, plus les variables sont décorréliées, plus il est proche de -1 ou 1, plus elles sont corrélées. Le signe indique si les deux variables sont liées par une fonction croissante ou décroissante.

$$Cor(X, Y) = \frac{Cov(X, Y)}{\sigma_X \sigma_Y} \quad (3.13)$$

avec X et Y deux variables, $Cov(X, Y)$ leur covariance et σ_X et σ_Y leurs écarts-types.

$Cor(X, Y)$ est aussi appelé *coefficient de corrélation linéaire*, noté r . Une formule plus détaillée est décrite en Équation 3.14.

$$r = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2} \sqrt{\sum_i (y_i - \bar{y})^2}} \quad (3.14)$$

avec $\bar{x}$ et $\bar{y}$ les moyennes des variables X et Y .

Une autre approche de la dépendance entre variable repose sur la mesure d'*information mutuelle*, qui découle de la théorie de l'information. L'information mutuelle est évaluée via l'Equation 3.16 et permet également de mesurer la corrélation entre deux variables.

$$I(X, Y) = \sum_{x \in X, y \in Y} P(X, Y) \times \log\left(\frac{P(X, Y)}{P(X) \times P(Y)}\right) \quad (3.15)$$

Yu et Liu utilisent également une mesure reposant sur la théorie de l'information en utilisant le concept de gain d'information présenté précédemment [Yu2003]. Ils utilisent le concept d'*incertitude symétrique*, définit par l'équation suivante :

$$SU(X, Y) = 2 \times \frac{IG(X, Y)}{H(X) + H(Y)} \quad (3.16)$$

avec $IG(X, Y)$ le gain d'information de X sachant Y , et $H(X)$ et $H(Y)$ les entropies des variables X et Y .

Ces valeurs sont comprises entre 0 et 1, 1 indiquant que l'information contenue par une des deux

variables permet de prédire complètement l'autre variable, 0 indiquant que X et Y sont indépendantes.

Après s'être intéressé aux mesures de dépendances, nous développons dans la prochaine section les principes des approches fondées sur les mesures de consistance pour évaluer la performance d'un sous-ensemble de descripteurs.

3.2.4.4 Mesures de consistance

Les mesures de consistances cherchent à évaluer la capacité d'un sous-ensemble de paramètres à discriminer plusieurs classes. Un paramètre est dit *consistant* si il contient l'information suffisante pour séparer les différentes classes. Dans le cas contraire, il est dit *non consistant*.

La Figure 3.7 illustre cette notion avec une situation de discrimination entre deux classes. L'attribut 2 est clairement consistant car il prend des valeurs bien distinctes pour les points de la classe 1 et de la classe 2, ce qui permet de bien les différencier. Ce n'est pas le cas pour l'attribut 1, dont les valeurs ne sont pas clairement différenciables entre la classe 1 et 2.

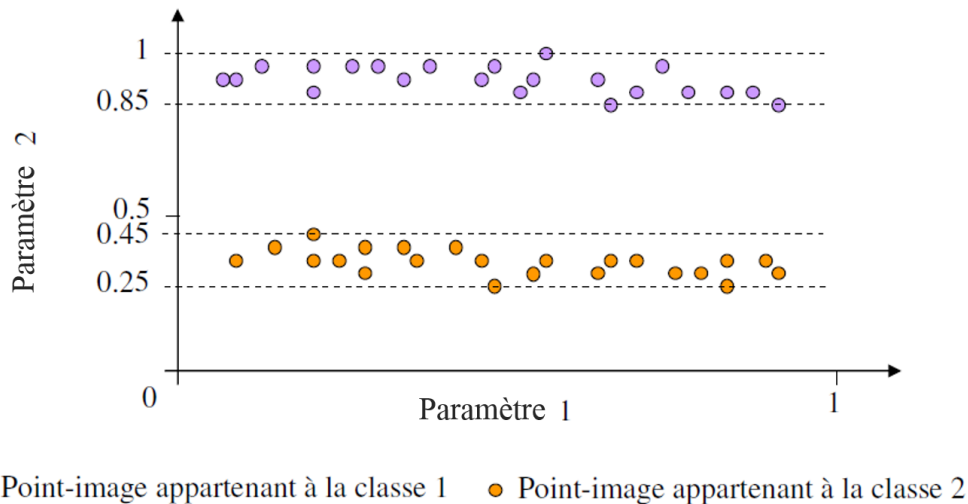


Figure 3.7 – Représentation d'un attribut consistant (paramètre 2) et d'un attribut non consistant (paramètre 1) [Porebski2009]

avec $P(X,Y)$ la probabilité conjointe de X et Y et $P(X)$ la probabilité d'occurrence de X

Si X et Y sont indépendants, donc non corrélés, $P(X,Y) = P(X) \times P(Y)$ et donc $I(X,Y) = 0$: l'information mutuelle est nulle.

Cette technique se démarque de la précédente, basée sur le coefficient de corrélation, dans le fait qu'elle permet de quantifier la corrélation entre deux variables, mais également leur dépendance statistique, ce qui la rend plus robuste.

Après s'être intéressé aux mesures de dépendances, nous proposons enfin d'exposer les principes des approches fondées sur les mesures directes du taux d'erreur de classification pour évaluer la performance d'un sous-ensemble de descripteurs.

3.2.4.5 Mesures du taux d'erreur de classification

Les mesures de taux d'erreur de classification, aussi appelées approches *wrapper* comme exposé au début de cette section, permettent d'évaluer les sous-ensembles de paramètres directement

en fonction de la performance de classification résultant de leur utilisation. Le sous-ensemble le plus pertinent est alors celui permettant d'obtenir un taux d'erreur de classification le plus faible sur un lot de données de test. Ce taux correspond généralement au ratio du nombre de données classées correctement et du nombre total de données testées.

Ce type de mesures permet d'être très précis dans le choix du sous-ensemble car celui-ci est choisi directement en fonction du problème de classification visé. Néanmoins, étant donné qu'il est nécessaire d'effectuer une classification avec chaque candidat, cette approche se révèle très coûteuse en temps de calcul, particulièrement lorsque :

- le nombre d'attributs initial est élevé,
- le nombre de données d'apprentissage et de test est élevé,
- les algorithmes de classification et de génération utilisés sont eux-même coûteux en temps de calcul.

Lors du processus de sélection de descripteurs, la plupart des méthodes mises en œuvre utilisent un critère d'arrêt pour stopper l'algorithme lorsqu'une solution satisfaisante est atteinte. Les différents types de critères communément utilisés sont présentés dans la section suivante.

3.2.5 Critère d'arrêt

Le critère d'arrêt tient un rôle très important dans les algorithmes de sélection de descripteurs, notamment pour les approches séquentielles et aléatoires qui utilisent des processus itératifs ne parcourant pas l'espace des solutions de manière complète. Les performances et la dimension des sous-espaces obtenus par ces techniques dépendent fortement du moment où le processus est stoppé. D'un autre côté, le critère d'arrêt permet également de garantir un temps de calcul raisonnable. Il s'agit alors de trouver un consensus entre performance, dimension du sous-espace et temps de calcul.

Pour répondre à cette problématique, deux grandes catégories de critères sont communément référencés :

- Critères **basés sur le processus de génération** : ce type de critère permet de stopper le processus soit lorsque le sous-espace solution est composé d'un certain nombre de descripteurs, soit lorsqu'un certain nombre d'itérations ont été réalisées.
Si ce type de critère est très simple à mettre en place, le choix du nombre de descripteurs ou d'itérations optimal est très compliqué à fixer *a priori*. Il s'agit en général d'une décision arbitraire qui peut mener à un arrêt prématuré ou au contraire trop tardif menant à des performances finales médiocres [Doak1992] [Porebski2009].
- Critères **basés sur la fonction d'évaluation** : ce type de critère permet de stopper le processus soit lorsqu'un certain seuil est atteint par la fonction d'évaluation, soit lorsque la valeur de la fonction d'évaluation n'évolue pas suffisamment d'une itération à une autre, c'est-à-dire lorsque la modification du sous-ensemble solution n'apporte pas un gain de discrimination significatif.
Ces méthodes permettent de s'assurer d'une certaine performance de la solution retenue au regard de la fonction d'évaluation choisie, contrairement au critères basés sur la génération qui donnent moins de garanties à ce sujet.

Une fois la sélection stoppée et un sous-espace de descripteurs retenu, il est nécessaire de valider cette solution afin de vérifier qu'elle donne en effet de bons résultats pour le problème de classification considéré. Un aperçu des principaux types de méthodes de validation est présenté dans la prochaine section.

3.2.6 Validation

L'étape de validation suit le processus de sélection et a pour objectif de tester la pertinence du sous-espace solution retenu. Les données disponibles sont souvent séparées en une partie utilisée pour l'entraînement de l'algorithme de classification et pour le tri des descripteurs, appelées *données d'apprentissage*, et une partie utilisée pour réaliser la validation, appelées *données de test*. L'erreur de classification pour l'ensemble de test est généralement calculée et utilisée comme indicateur de la performance de la solution testée. Cette erreur est liée à l'utilisation d'un algorithme de classification, dont les principales familles sont présentées dans le chapitre 4 de ce mémoire. Plusieurs approches peuvent être considérées lorsqu'il s'agit de former le groupe des données de test :

- La validation par **resubstitution** : elle consiste en l'utilisation du même ensemble de données pour l'apprentissage et le test. Cette méthode permet d'avoir une idée globale du fonctionnement ou non de la solution retenue, mais elle sous-estime souvent l'erreur de classification. En effet, ce sont les mêmes données qui servent à la fois à entraîner et à valider le modèle, il est donc souvent plus intéressant de tester le modèle sur des données nouvelles par rapport à l'entraînement.
- La validation par **validation croisée** ou **cross-validation** en anglais : elle regroupe plusieurs méthodes qui séparent les données en un groupe dédié à l'apprentissage et un groupe dédié au test. Plusieurs types de validation croisée existent :
 - la méthode *testset validation* ou *holdout validation* : l'ensemble des données est divisé en deux partitions, une pour l'apprentissage, l'autre pour la validation. La partition d'apprentissage contient communément au moins 50 à 75% des données et le reste est versé dans la partition de test. C'est la forme la plus simple de validation croisée.
 - la méthode *k-fold cross-validation* : l'ensemble des données est divisé en k échantillons, puis on réalise k apprentissages pour tester alternativement chacun d'eux en utilisant à chaque fois les $(k - 1)$ échantillons restants pour former l'ensemble d'apprentissage.
 - la méthode *leave-one-out-cross-validation* (LOOCV) : il s'agit d'un cas particulier de la méthode précédente pour $k = N$, N étant le nombre de données total. Ainsi, chaque donnée de l'ensemble initial est testée à tour de rôle individuellement face à un modèle appris sur toutes les autres données restantes. Elle est plutôt recommandée dans le cas d'ensembles de données de petite taille car elle devient très coûteuse en temps de calcul pour de grands ensembles.
 - la méthode *bootstrap* : l'ensemble d'apprentissage est constitué en tirant avec remise un certain nombre de données depuis l'ensemble total. Une donnée peut alors ne pas être présente du tout ou être présente plusieurs fois dans l'ensemble d'apprentissage. L'ensemble de test est ensuite constitué soit de l'ensemble total des données, soit des données non sélectionnées.

De manière générale, les méthodes de validation croisée sont préférées car elles permettent de donner une valeur d'erreur de classification plus proche de la réalité.

Nous avons donc présenté les principes et caractéristiques des méthodes de sélection de descripteurs. Ces techniques permettent de ne sélectionner, à partir d'un ensemble initial plus ou moins grand, seulement que les descripteurs les plus pertinents pour résoudre un problème donné. Pour faire suite à cette partie, nous souhaitons résumer les principes d'une autre famille de méthodes permettant de réduire l'information obtenue après la phase d'extraction de descripteurs depuis les images. Il s'agit des méthodes de réduction de dimensionnalité par transformation de l'espace des variables.

3.3 Transformation de l'espace des descripteurs

Les techniques de réduction de dimensionnalité par transformation de l'espace des descripteurs font parties de la famille des méthodes statistiques d'*analyse de données*, qui permettent d'étudier la structure d'un lot de données pour en dégager certains aspects intéressants. La réduction de dimensionnalité basée sur ces méthodes permet de transformer les variables, qui correspondent ici aux descripteurs extraits, selon un aspect statistique donné, afin d'obtenir un nouvel espace de variables de dimensions plus faibles.

Il existe un grand nombre de technique de réduction de dimensionnalité. Nous choisissons de présenter les principes de fonctionnement de 4 d'entre elles, parmi les plus utilisées :

- l'*Analyse en Composantes Principales* (ACP) qui cherche à maximiser la variance du lot de données tout en produisant un espace de nouvelles variables décorréliées entre elles ;
- l'*Analyse en Composantes Indépendantes* (ACI) qui cherche à produire un nouvel espace de variables les plus statistiquement indépendantes les unes des autres ;
- l'*Analyse Linéaire Discriminante* (LDA pour *Linear Discriminant Analysis*) qui cherche à créer un nouvel espace de variables permettant de séparer au mieux les classes présentes dans les données.
- l'*algorithme t-distributed Stochastic Neighbor Embedding* (t-SNE) qui cherche un espace de faible dimensions optimal qui respecte les proximités entre les données dans l'espace de plus haute dimension.

Les deux premières techniques sont considérées comme non-supervisées car elles ne tiennent pas compte de la catégorisation éventuelle des données en différentes classes. La variance ou l'indépendance des données sont évaluées à partir de l'information *intrinsèque* des données.

A l'inverse, la LDA est considérée comme supervisée car elle tient compte de l'information de catégorisation des données en différentes classes. La construction du nouvel espace s'appuie dans ce cas sur cette information *extrinsèque* aux données en elles-même.

La dernière méthode est également une approche non-supervisée mais se différenciant des 3 autres dans son aspect non-linéaire. Ce type d'approche permet de mieux représenter l'information de proximité entre les points pour la réduction d'espaces de descripteurs de grandes dimensions.

Une description des principes généraux de fonctionnement de chacune de ces méthodes est présentée ci-après. Nous ne souhaitons pas entrer ici dans le détail du fonctionnement de chaque méthode, mais plutôt donner un aperçu des objectifs de chacune d'elles. Néanmoins, nous invitons le lecteur à consulter la thèse de Tollari qui propose une description plus complète du fonctionnement de certaines de ces techniques [Tollari2006].

3.3.1 Analyse en Composantes Principales (ACP)

L'ACP est fondée sur des travaux réalisés par Pearson sur les corrélations entre variables [Pearson1901], puis a été développée et formalisée par la suite par Hotelling [Hotelling1933]. Il n'est d'ailleurs pas rare de la retrouver sous le nom de *transformée de Hotelling*, ou encore de *transformée de Karhunen-Loève*, du nom de deux autres chercheurs ayant également participé à sa formulation, suivant les domaines d'application considérés. Il s'agit d'une des méthodes les plus utilisées pour la réduction de dimensionnalité.

A partir d'un ensemble de descripteurs de départ, l'ACP permet de trouver un sous-espace de nouveaux descripteurs permettant de maximiser la variance des données considérées. Les nouveaux descripteurs sont construits comme des combinaisons linéaires des descripteurs initiaux. La Figure 3.8 permet de visualiser ce principe de maximisation de la variance sur un exemple très simple. Les vecteurs V_1 et V_2 sont appelés *composantes principales* ou *axes principaux* du nuage

de données décrites par les descripteurs D_1 et D_2 . La variance des données est maximale dans les directions de ces deux vecteurs. Dans ce cas simple, aucune réduction de dimensionnalité n'est réalisée car on a 2 descripteurs en sortie pour 2 en entrée.

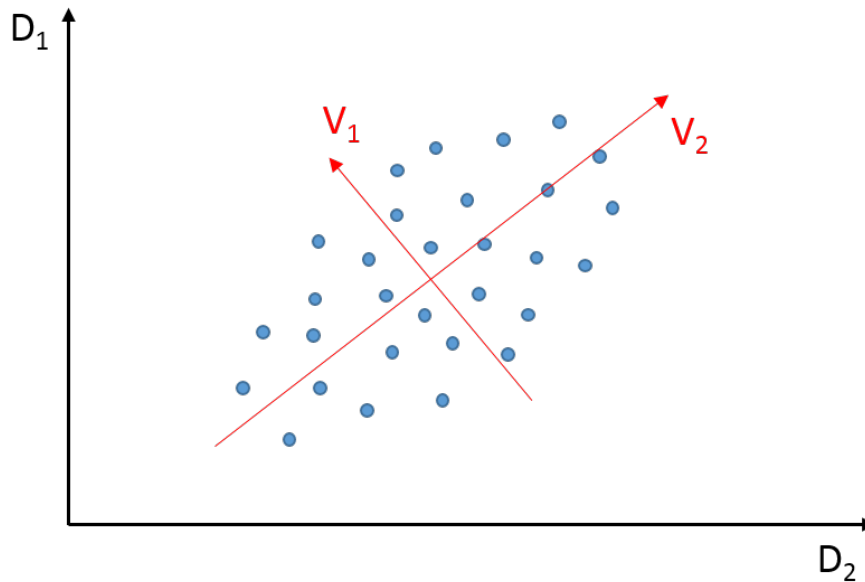


Figure 3.8 – Illustration des axes V_1 et V_2 choisis pour maximiser la variance dans la cadre d'une ACP sur un lot de données décrits au départ par les descripteur D_1 et D_2

Lorsqu'on souhaite réduire le nombre de dimensions, on choisit les n premiers axes principaux qui permettent d'expliquer une certaine proportion de la variance totale des données. L'opération engendre une perte irréversible d'information et l'objectif est d'obtenir un sous-espace qui permet de minimiser cette perte.

La Figure 3.9 illustre ce cas pour une ACP réalisée à partir de 14 descripteurs initiaux (B).

Le résultat de l'opération permet d'obtenir 14 axes propres (V) décorrélés, issus de combinaisons linéaires des 14 descripteurs originaux, dont les coefficients sont partiellement visibles dans le tableau (c). Chaque axe participe à l'explication de la variance totale à hauteur d'une certaine proportion exposée dans le tableau (b). Dans ce cas, on peut par exemple choisir de ne garder que les 3 premières composantes qui expliquent déjà 83% de la variance en cumulé. Cela permet donc de passer de 14 descripteurs initiaux à 3 nouveaux descripteurs non corrélés entre eux, tout en gardant une forte proportion de l'information en variance. La vue (a) permet d'illustrer la projection des données dans le nouvel espace à partir des 3 premières composantes. Les anciens axes de projection sont affichés au centre du cube de représentation.

La réduction de dimensionnalité via ACP a par exemple été utilisée par Malhi et Gao pour une application de surveillance de machine dans le cadre de la production de roulements à billes [Malhi2004]. Elle leur a permis d'obtenir des descripteurs plus pertinents que leur ensemble initial pour la classification de défauts sur les roulements.

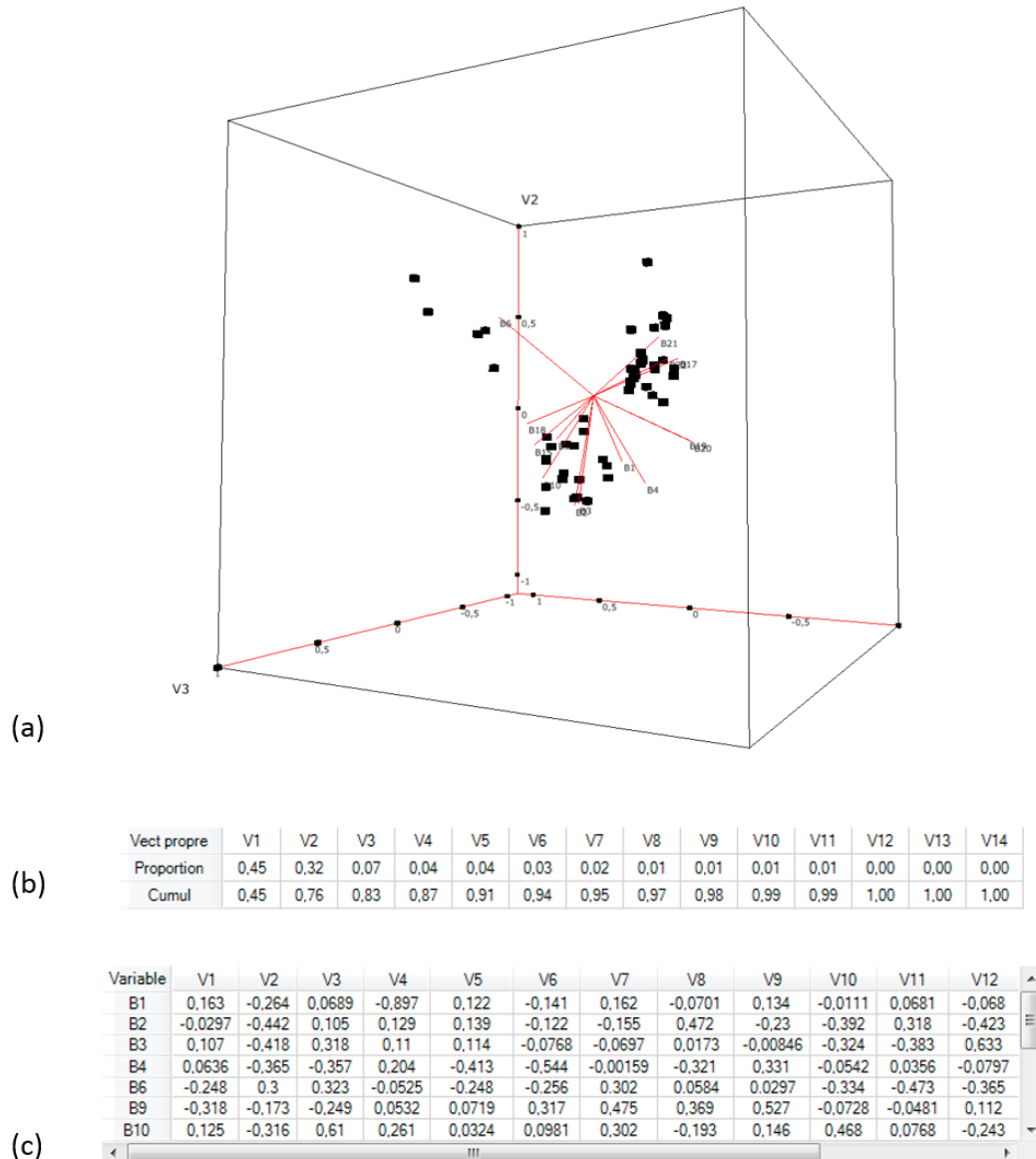


Figure 3.9 – Illustration d'une réduction de dimensionnalité via la méthode ACP - (a) Projection des données selon les 3 premières composantes principales - (b) Proportion de la variance expliquée par les composantes principales (vecteurs propres) - (c) Coordonnées des variables initiales dans l'espace de composantes principales (coefficients des combinaisons linéaires)

Si l'ACP permet d'obtenir des composantes principales décorrélées entre elles, elle ne garantit pas que ces nouvelles variables soit indépendantes entre elles. Afin de répondre à des problèmes qui cherchent à obtenir cette caractéristique, il est possible d'utiliser une autre méthode d'analyse statistique : l'analyse en composantes indépendantes.

3.3.2 Analyse en Composantes Indépendantes (ACI)

L'ACI est issue des travaux menés par Héroult et Jutten dans le domaine des neurosciences et du traitement du signal, notamment pour résoudre des problèmes de séparation aveugle de sources [Héroult1985]. Elle a plus tard été formalisée de manière mathématique par Comon [Comon1994].

Tout comme l'ACP, les composantes générées par l'ACI sont des combinaisons linéaires des

descripteurs initiaux, mais avec l'objectif de produire des variables les moins dépendantes possibles les unes des autres. Nous ne détaillons pas plus le fonctionnement de cette approche car elle est généralement moins utilisée que l'ACP pour des applications de réduction de dimensionnalité.

On peut néanmoins citer les travaux de Bressan qui utilise une représentation par ACI données de grandes dimensions pour améliorer les performances d'un algorithme de reconnaissance d'objets basé sur une classification bayésienne. Ils montrent notamment que la caractéristique d'indépendance des composantes obtenues permet d'améliorer la précision du classifieur utilisé ensuite.

L'ACP et l'ACI sont donc des méthodes permettant de réaliser une réduction de dimensionnalité de manière non supervisée, c'est à dire ne tenant pas compte d'un éventuel étiquetage des données suivant différentes classes. Cependant, on peut souhaiter dans certains cas réduire la dimensionnalité d'un ensemble tout en préservant au maximum les classes. Pour cela, il est possible d'utiliser d'autres types de méthodes d'analyse statistiques, telle que l'analyse linéaire discriminante.

3.3.3 Analyse Linéaire Discriminante (LDA)

La LDA tire ses sources des travaux de Ronald Fisher sur la caractérisation de fonctions discriminantes pour séparer des populations de données [Fisher1936].

Cette méthode consiste comme les deux précédentes, à rechercher des axes permettant de représenter les données en maximisant une de leurs caractéristiques. Dans le cas de la LDA, les axes sont recherchés tels que la séparation entre les différentes classes de données dans le nouvel espace de projection est maximale.

La Figure 3.10 expose une comparaison entre les axes principaux fournis par une analyse ACP et par une analyse LDA pour un même lot de données possédant 2 classes distinctives.

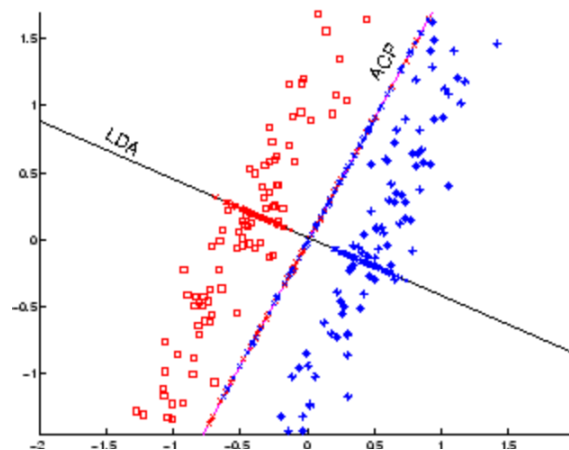


Figure 3.10 – Illustration du premier axe principal fourni par une ACP et une LDA sur un même ensemble de données

L'inconvénient connu de cette technique est qu'elle fait l'hypothèse que les classes suivent une distribution gaussienne, ce qui est rarement le cas dans des problèmes réalistes. Même si la LDA peut fonctionner de manière satisfaisante si cette hypothèse est légèrement compromise, il existe également d'autres méthodes de discrimination des classes ne faisant pas ce postulat. C'est le cas de la *Diversité Marginale Maximale* (MMD) par exemple, introduite plus récemment par Vasconcelos [Vasconcelos2003] pour réaliser de la réduction de dimensionnalité dans un contexte de reconnaissance visuelle.

Les trois méthodes présentées précédemment permettent une réduction linéaire d'un espace de descripteurs. Dans certains cas, il est néanmoins nécessaire d'utiliser des approches de réduction non-linéaires, qui tentent en général de mieux préserver l'information de proximité entre les données. Nous proposons pour terminer cette partie de présenter un représentant de ces approches non-linéaires : l'algorithme t-SNE.

3.3.4 Algorithme t-SNE

L'algorithme t-SNE est une méthode de réduction de dimensionnalité non-linéaire proposée récemment par Van Der Maaten et Hinton, particulièrement adaptée pour la réduction d'ensembles de grande dimensionnalité [Maaten2008]. Elle est une extension de la méthode du *Stochastic Neighbor Embedding* introduite antérieurement par Hinton et Roweis [Hinton2003].

Cette méthode permet de réduire la dimension d'un espace de descripteurs selon un critère probabiliste caractérisant la proximité entre les individus, suivant le principe de fonctionnement suivant :

1. Création de la *distribution de probabilité de l'espace de départ* grande dimension : la distance euclidienne (ou autre) entre les différentes données dans l'espace de grande dimension est tout d'abord convertie en une probabilité conditionnelle représentant les similarités entre les données. La distribution de cette probabilité est telle que des points proches l'un de l'autre dans l'espace de départ ont plus de chance d'être tirés que des points éloignés.
2. Création de la *distribution de probabilité de l'espace d'arrivée* (petite dimension) : une distribution similaire à la première est définie pour l'ensemble de faible dimension.
3. Minimisation de la *divergence entre les deux distributions* : la divergence de Kullback-Leibler¹ entre les deux distributions est enfin minimisée par rapport à l'emplacement des données dans l'espace de départ, ce qui permet d'obtenir au final les composantes de l'espace d'arrivée.

Pour plus d'informations quant au fonctionnement détaillé de l'algorithme, nous invitons le lecteur à consulter l'article de Van Der Maaten et Hinton qui présente cette méthode [Maaten2008].

Pour conclure sur les approches et méthodes existantes pour la réduction de dimensionnalité en parallèle de celles de sélection de descripteurs, nous proposons une mise en perspective rapide de ces concepts dans le cadre du contrôle visuel automatisé en industrie.

3.3.5 Synthèse et mise en perspective dans la cadre du contrôle visuel automatisé en industrie

Les méthodes de réduction de dimensions par transformation de l'espace des descripteurs permettent donc de générer de nouveaux attributs en cherchant à maximiser une caractéristique statistique des données. L'ACP cherche à maximiser la variance des données, l'ACI l'indépendance entre les variables, la LDA la discrimination entre les classes et le t-SNE à préserver la proximité entre les données. Le Tableau 3.3 résume les principales caractéristiques de ces 4 méthodes.

Cependant, dans le cadre d'une application dans un contexte de contrôle sur des lignes de production industrielles, nous avons mis en avant le fait que le système développé doit pouvoir fonctionner avec des temps de calcul courts. Comme le fait remarquer Porebski, ce critère constitue une limitation pour l'utilisation des méthodes de transformation de l'espace dans ce contexte

1. La divergence de Kullback-Leibler est une mesure de dissimilarité entre deux distributions de probabilités.

Méthode	ACP	ACI	LDA	t-SNE
Rapport aux classes de données	Non supervisé	Non supervisé	Supervisé	Non supervisé
Approche de transformation	Linéaire	Linéaire	Linéaire	Non linéaire
Information privilégiée	Variance des données	Indépendance des descripteurs	Discrimination entre les classes	Proximité des données

Tableau 3.3 – Principales caractéristiques des 4 méthodes présentées

de classification en-ligne [Porebski2009] . En effet, l'inconvénient de ce type de techniques est qu'elles nécessitent de recalculer la projection des données extraites des images dans le nouvel espace à chaque contrôle.

Ce n'est pas le cas des méthodes de sélection de paramètres, qui permettent même au contraire de diminuer le temps de calcul en-ligne car seuls les descripteurs retenus au préalable sont extraits à chaque contrôle. Aucun calcul supplémentaire n'est alors nécessaire.

A nombre égal de descripteurs finaux, les méthodes de sélection sont donc plus rapides pour réduire l'information et il est donc préférable de les utiliser dans un contexte de contrôle industriel.

Dans la seconde partie de ce chapitre, nous souhaitons présenter un critère d'évaluation de la discrimination entre deux classes, que nous avons développé au cours de ce travail de thèse. Ce critère est fondé sur une analyse multivariée des données via le test du T^2 .

3.4 Critère de séparation basé sur le test du T^2

Comme exposé dans la partie précédente, il existe différents types de fonctions d'évaluation pour les méthodes de sélection de descripteurs. Dans le cadre de ce travail de thèse, nous nous sommes plus spécifiquement intéressés aux méthodes de mesures de distance, notamment au travers de l'utilisation du test statistique du T^2 .

Dans cette partie, nous proposons tout d'abord de développer les fondements de ce test, ainsi qu'une application à la sélection de descripteurs qui en est faite dans la littérature. Puis, partant de ces considérations, nous exposerons le principe de fonctionnement d'un nouveau critère d'évaluation proposé dans ce travail de thèse.

3.4.1 Test du T^2

Le test statistique du T^2 , aussi appelé T^2 de Hotelling, est issu des travaux de Hotelling sur les distributions multivariées [Hotelling1931]. Il s'agit d'une généralisation du populaire test t de Student qui avait été introduit au préalable par Gosset ².

2. W.L. Gosset était employé par l'entreprise Guinness lorsqu'il inventa ce test. La politique de l'entreprise étant de ne pas publier sur les découvertes de ces chercheurs, Gosset fut seulement autorisé à publier sous un pseudonyme, *Student*. Le fameux test a donc été nommé après ce pseudonyme et non selon le véritable nom de son inventeur.

Ces tests peuvent par exemple permettre de comparer deux populations de données suivant certaines caractéristiques statistiques pour indiquer si elles sont différentes l'une de l'autre. Ils peuvent également permettre de savoir dans quelle mesure les différences sont significatives, et donc en d'autres termes si les différences observées entre les données sont seulement dues au hasard ou non.

Le test t ne s'applique qu'à des distributions univariées, c'est-à-dire à des données décrites par une seule variable, alors que le test du T² permet une généralisation aux distributions multivariées, c'est-à-dire à des données décrites par plusieurs variables.

3.4.1.1 Principes du test statistique

Les méthodes de test statistiques utilisent une fonction nommée *statistique de test*, qui permet de quantifier le comportement des observations pour pouvoir répondre au test considéré. Ce test statistique consiste en une procédure de décision entre deux hypothèses. D'un côté une *hypothèse nulle* et d'un autre côté une *hypothèse alternative*. Dans le cadre de la comparaison de deux populations par exemple, l'hypothèse nulle postule une égalité entre les paramètres statistiques des deux populations, et l'hypothèse alternative postule une différence significative entre les deux populations. Il s'agit alors de rejeter ou ne pas rejeter l'hypothèse nulle en fonction d'un jeu de données, et d'un risque α choisi.

Nous ne développons pas plus ce concept de test ici, mais pour aller plus loin, nous invitons le lecteur à visiter la [page wikipédia du test de Student](#) où un exemple d'utilisation du test de Student est présenté.

Nous nous intéressons dans la suite aux statistiques de test et à leur application à la sélection de descripteurs.

3.4.1.2 Statistique de test du T²

Dans le cadre du test t de Student sur un échantillon, la statistique de test pour la comparaison de l'échantillon y de taille n à une population hypothétique de moyenne μ est définie de la manière suivante :

$$t = \frac{\bar{y} - \mu}{s} \sqrt{n} \quad (3.17)$$

avec s l'écart-type de l'échantillon :

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2} \quad (3.18)$$

Si on élève cette statistique au carré on obtient alors :

$$t^2 = \frac{(\bar{y} - \mu)^2}{s^2} n = n(\bar{y} - \mu)(s^2)^{-1}(\bar{y} - \mu) \quad (3.19)$$

Dans le cadre de la généralisation de Hotelling à des données multivariées, l'échantillon y unidimensionnel devient un échantillon Y composé de n composantes-vecteurs indépendants de dimension p :

$$Y = \begin{bmatrix} y_{11} & y_{12} & \dots & y_{1n} \\ y_{21} & y_{22} & \dots & y_{2n} \\ \dots & \dots & \dots & \dots \\ y_{p1} & y_{p2} & \dots & y_{pn} \end{bmatrix} = [Y_1, Y_2, \dots, Y_n] \quad (3.20)$$

La statistique définie en équation 3.21 devient alors :

$$T^2 = n(\bar{Y} - \mu)S^{-1}(\bar{Y} - \mu) \quad (3.21)$$

avec $\bar{Y}$ le vecteur des moyennes de Y et S la matrice de covariance de dimension $p \times p$ estimée sur l'échantillon Y .

Cette définition peut être étendue à un test du T^2 non plus à un échantillon, mais à deux échantillons. L'hypothèse nulle est alors formulée comme l'égalité entre les moyennes des deux échantillons. Ce test du T^2 à deux échantillons a été utilisé récemment par Porwik pour sélectionner des descripteurs dans le cadre d'application de reconnaissance de signatures manuscrites [Porwik2015].

A partir de cette brève présentation des principes de la statistique du T^2 , nous souhaitons introduire son utilisation pour des applications de cartes de contrôle.

3.4.2 Application aux cartes de contrôle

Une application de la statistique du T^2 de Hotelling réside dans la mise au point de cartes de contrôle dans le domaine du contrôle qualité pour la maîtrise de processus. Ce type de carte permet de détecter l'apparition d'une variation de caractéristique entraînant une altération du processus contrôlé.

La particularité des cartes du T^2 est qu'elles permettent de contrôler plusieurs caractéristiques en même temps grâce à l'aspect multivarié de la statistique du T^2 .

De manière simplifiée, le calcul de ce type de carte suit les étapes suivantes :

1. Un ensemble de mesures est réalisé sur le processus, permettant de récupérer des valeurs pour chaque caractéristique (variable) identifiée pour décrire le processus.
2. La loi de probabilité (considérée comme Gaussienne) suivie par le processus est caractérisée en calculant les moyennes (vecteur μ) et la covariance (Σ) des caractéristiques mesurées, suivant deux approches possibles :
 - Calcul par rapport aux moyennes et à la covariance des caractéristiques de la population entière des observations.
 - Calcul par rapport aux moyennes et à la covariance des caractéristiques d'un échantillon identifié d'observations.
3. Pour chaque individu X de l'échantillon et par rapport à la loi de probabilité précédente, une distance est calculée à partir de la statistique du T^2 :

$$T^2 = (X - \mu)\Sigma^{-1}(X - \mu) \quad (3.22)$$

Il est possible de remarquer que cette distance se rapproche de celle proposée par Mahalanobis [Mahalanobis1936], introduite précédemment dans les mesures de distance pour la sélection de donnée :

$$d(x, y) = \sqrt{(X - \mu)^T \Sigma^{-1} (X - \mu)} \quad (3.23)$$

4. Un seuil de contrôle est fixé au travers du choix de la valeur du risque α . Il permet de caractériser la limite au delà de laquelle on considère qu'un individu est significativement différent de la population de référence.

La Figure 3.11 illustre une carte de contrôle pour un processus multivarié à 5 variables. Les moyennes et la covariance de ces variables ont été calculées sur un échantillon d'apprentissage de 10 données. Le risque α qui permet de fixer le seuil de contrôle a été fixé à 0,5%. Sous ces conditions, on peut remarquer que 4 individus dans l'échantillon de test sont significativement différents du reste des données.

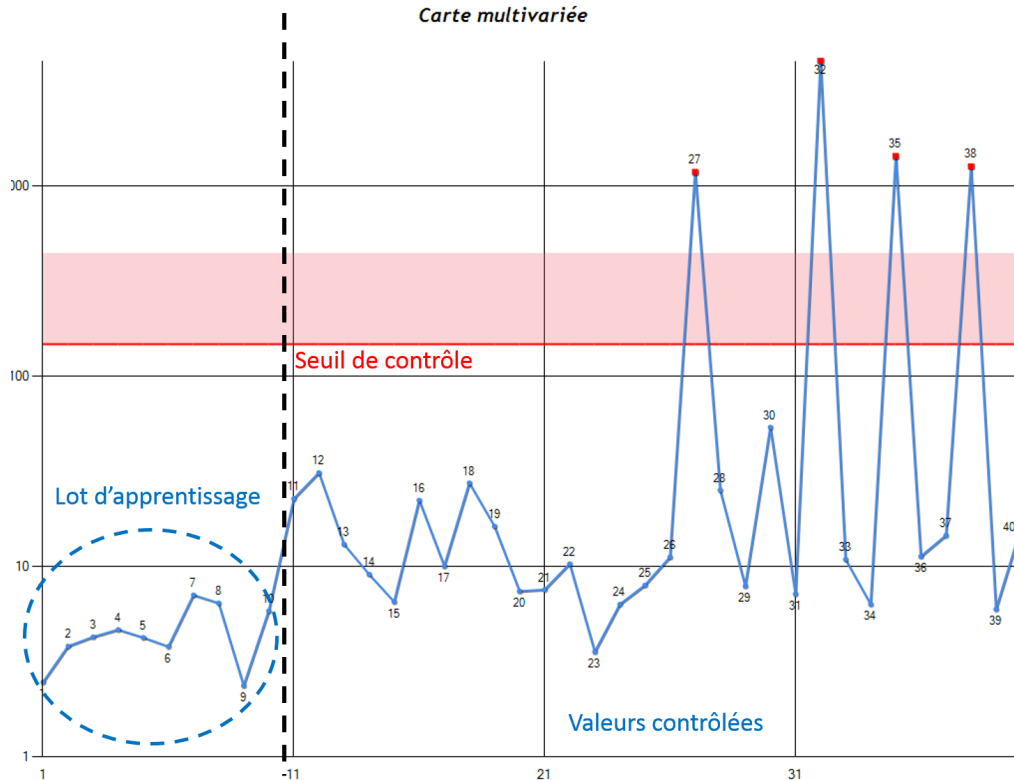


Figure 3.11 – Exemple d'une carte de contrôle du T^2 pour un processus multivarié à 5 variables. 10 observations servent à apprendre les moyennes et la covariance des variables du processus.

Si ce procédé a l'avantage de permettre de contrôler des données multivariées, il est important de garder en tête qu'il est néanmoins nécessaire de posséder plus de données que de variables pour pouvoir calculer la matrice de covariance, et donc pour calculer la distance utilisée.

Dans le cadre de la sélection de paramètres pour optimiser la discrimination entre 2 classes, nous proposons un critère d'évaluation inspiré des principes des cartes de contrôle présentés précédemment.

3.4.3 Critère d'évaluation de distance par le T^2

Dans le cadre de ces travaux de thèse, nous proposons une méthode d'évaluation de la discrimination entre deux classes grâce aux outils présentés précédemment. Cette méthode s'inscrit comme une fonction d'évaluation pouvant être utilisée pour réaliser une sélection de descripteurs.

Le principe de cette approche est d'utiliser une carte de contrôle calculée sur le sous-ensemble de descripteurs testé pour évaluer la séparation entre les individus de chaque classe via les descripteurs choisis.

En apprenant les paramètres statistiques (moyenne et covariance) d'un échantillon de données de classe C_1 , il est possible de définir une carte de contrôle permettant de caractériser la similarité

entre cet échantillon et des nouveaux individus test. Si les données de test contiennent des individus issus de la classe C_1 et d'une deuxième classe C_2 , il est possible d'évaluer grâce à la carte si l'information contenue dans le sous-ensemble de descripteurs est pertinente pour séparer les classes.

Le fonctionnement de cette méthode peut être décomposé en différentes étapes :

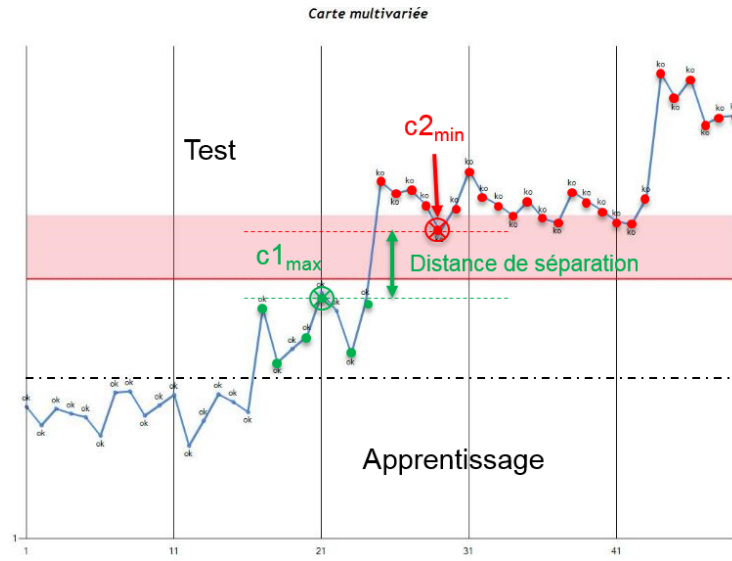
1. **Séparer les données d'entraînement** en deux groupes : un groupe dit d'*apprentissage*, correspondant à une classe C_1 , et un groupe dit de *test*³, correspondant à une classe C_2 . Il est important que le groupe d'*apprentissage* soit composé uniquement d'individus n'appartenant qu'à la classe C_1 , et que le groupe de *test* soit composé d'éléments des deux classes C_1 et C_2 .
2. **Évaluation de paramètres de la loi de probabilité** de la population de données d'apprentissage : les moyennes des différents descripteurs ainsi que la matrice de covariance de ces derniers sont calculées à partir des données d'apprentissage. Ils servent de référence pour le calcul de la prochaine étape.
3. **Génération de la "carte de contrôle"** : les distances du T^2 entre chaque individu du groupe test et le groupe de données d'apprentissage sont calculées. Le report dans un graphique des distances correspondant à chaque élément de test permet de générer la carte de contrôle.
4. **Calcul de la distance de séparation** : la différence entre la distance du point le plus "haut" de la classe C_1 de l'échantillon de test, nommé $c1_{max}$, et le point le plus "bas" de la classe C_2 de ce même échantillon, nommé $c2_{min}$ est appelée *distance de séparation*. Cette distance est maximale pour deux groupes de données très éloignés.

La Figure 3.12 illustre deux cas particuliers décrits par cette distance :

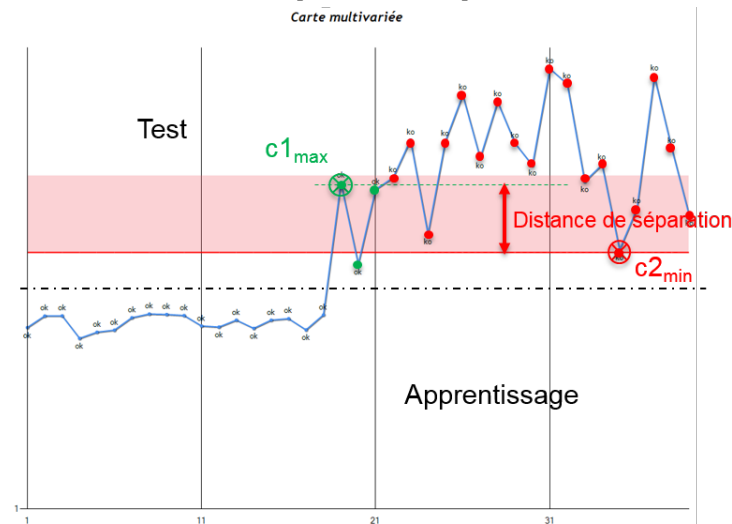
- en 3.12a, $c1_{max}$ est supérieur à $c2_{min}$, donc la distance de séparation est positive. Les deux échantillons sont donc séparés.
- en 3.12b, $c1_{max}$ est inférieur à $c2_{min}$, donc la distance de séparation est négative. Les deux échantillons ne sont donc pas séparés.

Du à l'utilisation de la matrice de covariance, cette méthode peut néanmoins être difficile à appliquer pour des données de haute dimensionnalité car elle nécessite plus d'observations (données) que de variables (descripteurs). Elle peut cependant être utilisée après une première réduction de l'ensemble des descripteurs si le nombre d'observation est insuffisant au départ.

3. Le groupe *test* dont nous parlons ici n'est pas le même que celui utilisé dans la phase de validation d'un sous-ensemble sélectionné. Il s'agit d'un groupe de données *test* faisant partie des données utilisées pour effectuer la sélection des paramètres.



(a) Exemple de bonne séparation



(b) Exemple de mauvaise séparation

Figure 3.12 – Des cas d'évaluation de la séparation entre des données de classe C_1 (vert) et C_2 (rouge). L'individu $c1_{max}$ est repéré par la croix verte et le $c2_{min}$ par une croix rouge.

Cette approche constitue donc un apport de ce travail de thèse concernant l'évaluation des performances de discrimination d'un sous-ensemble de descripteurs. Nous avons en premier lieu choisi de développer cette approche de part l'expertise présente au laboratoire SYMME sur l'utilisation de la statistique du T^2 . Cette approche ayant rapidement montré de bons résultats, nous l'avons utilisé pour le reste des développements décrits dans ce manuscrit. Consécutivement à cela, nous avons au final manqué de temps pour mener une étude comparative la confrontant aux autres approches utilisées dans la littérature et décrites dans la première partie de ce chapitre. Cette étude fait partie des perspectives identifiées pour la suite de ces travaux.

3.5 Conclusion

Nous avons donc exploré dans ce chapitre les différentes approches permettant de trier les descripteurs extraits des images afin de ne retenir que les plus pertinents pour l'étape suivante de classification. Cette étape est essentielle pour réduire l'ensemble des informations afin de garantir une classification performante et efficace en terme de temps de calcul.

Deux grandes familles se dégagent de la littérature :

- les approches par sélection de descripteurs qui consistent à réduire l'ensemble des attributs en choisissant un sous-espace de taille minimal permettant de répondre au mieux au problème de classification donné ;
- les approches par transformation de l'espace des descripteurs qui consistent en l'utilisation de méthodes généralement issues de l'analyse statistique de données, pour convertir l'espace initial des descripteurs en un espace de dimensionnalité plus faible tout en préservant une information spécifique.

Les approches de sélection de descripteurs sont en générale représentées comme la combinaison de 4 éléments :

- Une procédure de génération qui regroupe les approches complète ou exhaustive, séquentielle et aléatoire ;
- Une fonction d'évaluation qui peut être réalisée via des mesures de distances, d'information, de dépendance, de consistance ou directement du taux d'erreur de classification ;
- Un critère d'arrêt qui peut être basé soit sur le processus de génération, soit directement sur la fonction d'évaluation ;
- Une phase de validation du sous-espace retenu qui peut être menée soit grâce à une approche par resubstitution, soit grâce à une approche par validation croisée (cross-validation).

Si les approches de sélection fondées sur une procédure de génération complète et une évaluation directe du taux d'erreur de classification sont plus performantes que les autres approches, elles requièrent en général un temps de calcul important qui peut se révéler être un désavantage dans le contexte d'une application industrielle. Les approches de générations séquentielle ou aléatoire couplées à des mesures indirectes de la performance peuvent néanmoins apporter des résultats tout à fait satisfaisants tout en assurant un temps de calcul plus faible.

Les approches par transformation de l'espace des descripteurs regroupent un certain nombre de techniques se focalisant sur différents type d'information à préserver. Parmi les plus utilisées, nous avons cité l'ACP qui permet de maximiser la variance des données, l'ACI qui se concentre sur l'indépendance des descripteurs entre eux, la LDA qui cherche à maximiser la séparation entre les classes et le t-SNE qui se focalise sur la proximité entre les données.

Certaines de ces approches sont dites non-supervisées car elles étudient les données sans tenir compte d'une information de classe. Elles peuvent permettre d'identifier des caractéristiques de la structure des données non connues *a priori*. Cela peut notamment présenter un intérêt dans le cadre des approches de classification non-supervisées, abordées brièvement dans le Chapitre 4 de ce manuscrit.

Néanmoins, ce type d'approches présente un inconvénient important dans le cadre d'une application de contrôle en temps réelle. En effet, les données extraites des images nécessitent d'être projetées dans le nouvel espace pour chaque nouvelle classification, ce qui allonge le temps de calcul. Ce phénomène n'apparaît pas avec les approches de sélection, ce qui pousse à les préférer pour ce genre d'application.

Pour terminer, nous avons présenté les fondements de la statistique du T^2 d'Hotelling et des cartes de contrôles en maîtrise de processus, afin d'introduire un nouveau critère d'évaluation pouvant être utilisé pour la sélection de descripteurs. Cette approche s'applique à la séparation entre

deux classes de données et sera illustrée dans le dernier chapitre de ce manuscrit.

Une fois les images décrites grâce à des méthodes présentées précédemment dans le chapitre 2, puis l'ensemble des descripteurs extraits réduit grâce aux approches présentées dans ce chapitre, il est enfin possible d'utiliser les données ainsi préparées pour nourrir un algorithme de classification. Cette dernière étape permet tout d'abord d'apprendre les caractéristiques des différentes classes de données et ensuite d'attribuer une de ces classes à une nouvelle donnée (un produit à contrôler dans notre cas d'application). Le chapitre suivant présente les différentes approches de classification possible pour réaliser cette étape finale.

4

Classification de données

Sommaire

4.1	Introduction	115
4.2	Méthodes pour la classification de données	117
4.2.1	k-plus-proches voisins	117
4.2.1.1	Principe de fonctionnement	117
4.2.1.2	Distance entre deux éléments	118
4.2.1.3	Choix de k	118
4.2.1.4	Forces et limites	119
4.2.2	Réseaux bayésiens naïfs	120
4.2.2.1	Théorème de Bayes	120
4.2.2.2	Réseaux bayésiens	121
4.2.2.3	Classification par réseaux bayésiens naïfs	122
4.2.2.4	Forces et limites	123
4.2.3	Machines à Vecteurs de Support - Support Vector Machine (SVM)	124
4.2.3.1	Cas d'une séparation linéaire	124
4.2.3.2	Cas d'une séparation non-linéaire	125
4.2.3.3	Cas d'une classification multi-classes	127
4.2.3.4	Forces et limites	128
4.2.4	Arbres décisionnels et Forêts d'arbres décisionnels	128
4.2.4.1	Classification par arbre de décision	128
4.2.4.2	Forces et limites	132
4.2.4.3	Quelques évolutions des algorithmes d'arbres décisionnels	133
4.2.5	Réseaux de neurones	135
4.2.5.1	Neurone artificiel ou formel	135
4.2.5.2	Perceptron simple	137
4.2.5.3	Perceptron multi-couche	141
4.2.5.4	Forces et limites	144
4.2.6	Algorithme des k-moyennes	145
4.2.6.1	Principe de fonctionnement	145
4.2.6.2	Forces et limites	147
4.2.7	Réseaux de Deep Learning	147
4.2.7.1	Fondements historiques	147

4.2.7.2	Performances d'un réseau de Deep Learning	149
4.2.7.3	Utilisation de réseaux pré-entraînés	150
4.2.7.4	Forces et limites	151
4.3	Discussion quant aux caractéristiques des différentes méthodes et leur ap- plication industrielle	152
4.3.1	Comparaison des techniques	153
4.3.2	Mise en perspective dans le contexte du contrôle d'aspect industriel automatisé	157
4.3.3	Méthodes d'ensemble	160
4.3.4	Intérêt des algorithmes non-supervisés	160
4.4	Conclusion générale sur les méthodes de classification	162

4.1 Introduction

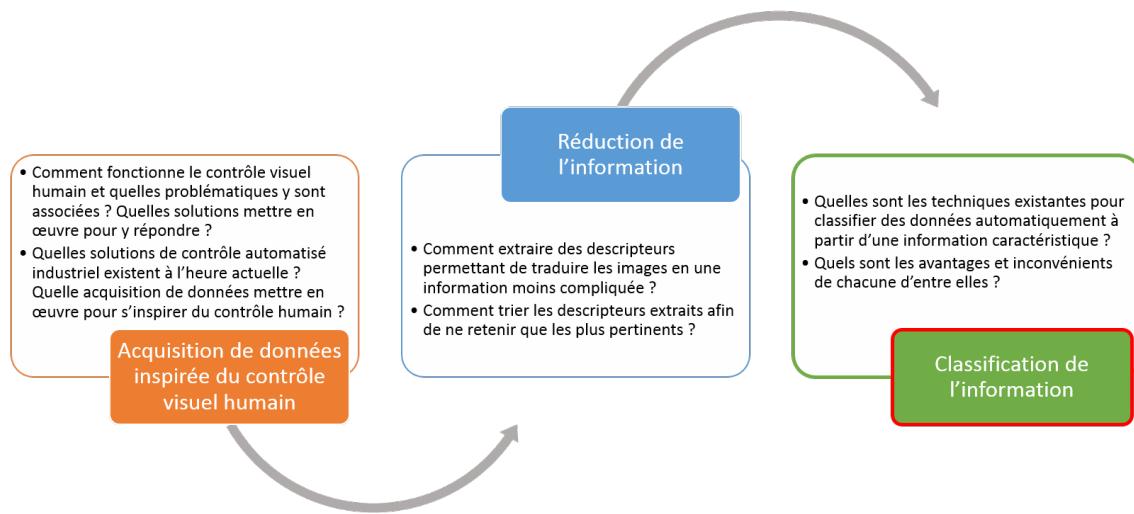


Figure 4.1 – Problématiques liées à la classification des données

Après réduction de l'information pour ne garder que des descripteurs pertinents à la description des images, ces derniers peuvent être utilisés afin d'attribuer des classes aux produits observés.

Les algorithmes de **classification automatique de données** cherchent à attribuer une *catégorie* ou *classe* à des *individus* ou *objets*, généralement à partir de modèles statistiques. Deux phases sont à distinguer lors de l'exécution de ces algorithmes : l'apprentissage et la classification en elle-même.

L'apprentissage consiste en l'entraînement de l'algorithme pour créer un modèle permettant de classer au mieux un ensemble de données d'entraînement. Il est possible de distinguer trois catégories de techniques d'apprentissage :

- **Les approches supervisées** : le système apprend à classer suivant un modèle prédéterminé par des données étiquetées fournies par un expert lors de la phase dite "hors ligne". Le système donne ensuite par lui-même des étiquettes aux données testées "en ligne" via le modèle préalablement établi.
- **Les approches non-supervisées** : le système est capable de créer lui-même un modèle de classes à partir d'un lot de données exemples non étiquetées. Avec ce type d'approches, aucun expert n'intervient à contrario des méthodes supervisées.
- **Les approches semi-supervisées** : le système est capable d'apprendre le modèle de classification à partir de quelques exemples étiquetés complétés par des exemples non-étiquetés. L'intervention de l'expert est alors réduite. Ces approches peuvent par exemple présenter un intérêt lorsque les lots de données en apprentissage sont grands et qu'il est fastidieux d'étiqueter tous les éléments.

Dans le cadre de notre travail, nous nous sommes particulièrement intéressés à l'application de techniques supervisées, pour garder l'avis de l'expert humain comme référence. Mais l'utilisation

de techniques non supervisées est également intéressante dans le cadre de la détection de nouvelles anomalies.

De nombreux algorithmes de classification supervisée et non-supervisée existent et il serait long de tous les expliciter. Nous souhaitons donc présenter dans cette première partie les principes de fonctionnement des techniques ou groupes de techniques les plus communément utilisées afin de fournir au lecteur une compréhension des mécanismes mis en jeu.

Les forces et limites seront exposés pour chaque technique puis synthétisées sous forme d'une comparaison des différentes méthodes dans la suite du chapitre. Les avantages et inconvénients de chaque méthode seront alors discutés, avant d'être mis en perspective dans le contexte du contrôle d'aspect industriel automatisé.

De manière générale, ce chapitre cherche donc à répondre aux questions : "Quelles sont les techniques existantes pour classifier des données automatiquement à partir d'une information caractéristique ?" et "Quels sont les avantages et inconvénients de chacune d'entre elles ?".

4.2 Méthodes pour la classification de données

Pour commencer, nous proposons donc de présenter les principales méthodes de classification de données utilisées en apprentissage automatique. Nous décrivons les principes de fonctionnement pour chaque famille de techniques, avant de présenter les forces et limites de chacune.

La plupart des techniques présentées font parties de la catégorie de l'apprentissage supervisé. En effet, nous souhaitons dans un premier temps partir d'exemples étiquetés donnés par un contrôleur humain. Parmi ces techniques sont présentées les k -plus-proches voisins, les réseaux bayésiens naïfs, les machines à vecteurs de support, les arbres décisionnels et les réseaux de neurones classiques et profonds. Nous choisissons également de présenter brièvement les k -moyennes en tant que technique non-supervisée. Elle peut en effet apporter un intérêt pour la détection de la structuration d'un ensemble de données ainsi que de nouvelles anomalies.

4.2.1 k -plus-proches voisins

La méthode des *k -plus-proches voisins* ou *k -NN* pour *k -nearest neighbors* est une méthode de classification supervisée introduite par Cover & Hart en 1961 [Cover1967]. C'est une des techniques les plus simples et intuitives de classification.

Elle fait partie des algorithmes dits *paresseux* (*lazy learning* en anglais), ce qui signifie qu'il n'y a pas de phase d'apprentissage explicite d'un modèle générique avant la classification, ou qu'elle est très minime. A contrario, la plupart des autres algorithmes présentés par la suite font partis des algorithmes dits *avides*¹ (*eager learning* en anglais), c'est-à-dire qu'ils cherchent à créer un modèle général à partir d'un ensemble d'entraînement avant de classer de nouvelles données.

4.2.1.1 Principe de fonctionnement

Le principe sous-jacent de la méthode des k -NN est que des éléments d'un lot de données se regroupent par sous-groupes d'éléments qui présentent des similarités. En d'autres termes, il existe généralement une certaine proximité entre éléments similaires dans l'espace de projection des données.

Le principe de fonctionnement de la méthode est alors de définir la classe d'une nouvelle donnée en fonction de celle des éléments qui sont ses plus proches voisins dans l'ensemble d'apprentissage. Afin de simplifier la visualisation du problème, on considère les données comme un ensemble de points dans un espace à n dimensions, définies par n descripteurs utilisées pour décrire les données. La classe majoritaire parmi les k plus proches points autour du nouvel élément à classer l'emporte, comme exposé en Figure 4.2 dans un cas simple à 2 dimensions. Comme indiqué précédemment, il n'y a donc pas de phase d'apprentissage d'un modèle avant la classification d'un nouvel élément, sa classe étant directement calculée à partir de la projection de données étiquetées dans un espace de n dimensions.

1. A ne pas confondre avec les algorithmes dits *gloutons* dont le principe est d'essayer d'atteindre un minimum global dans un problème d'optimisation en fixant un minimum local à chaque itération.

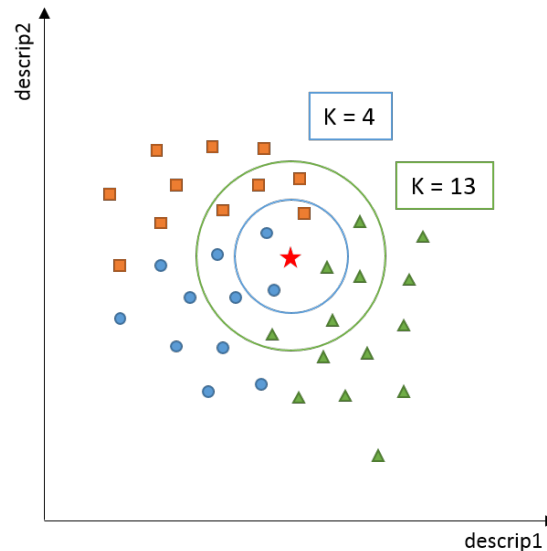


Figure 4.2 – Principe de fonctionnement de la méthode k -NN pour assigner une classe à une nouvelle donnée, dans un espace de projection composé de 2 descripteurs

Même si cette technique ne requiert pas de phase de construction de modèle pour classer les données, il est tout de même nécessaire de faire plusieurs choix avant de se lancer dans la classification de nouveaux éléments.

Tout d’abord, afin de caractériser la proximité entre les différentes données, il est nécessaire de choisir une métrique, appelée *distance*. Puis, il est également important de choisir une valeur de k permettant de déterminer au mieux la classe d’un nouvel élément. Ces deux problématiques sont brièvement décrites dans la suite de cette section.

4.2.1.2 Distance entre deux éléments

Le rôle du calcul de distance est de déterminer la position relative entre deux points. Les différentes distances qui peuvent être calculées sont les mêmes que celles présentées dans le Tableau 3.2 du Chapitre 3. Idéalement, le choix de la distance utilisée doit permettre de minimiser l’écart entre deux éléments d’une même classe, tout en maximisant l’écart entre deux éléments de classes différentes. La qualité de la classification par les k -plus-proches voisins dépend beaucoup du choix de cette distance.

Afin de diminuer la sensibilité des performances de l’algorithme au choix du calcul de la distance, plusieurs méthodes ont été proposées, dont une des plus notables est l’utilisation de système de pondération. L’objectif de ces techniques est de paramétrer le calcul de distance avec des poids appliquées aux descripteurs des données, afin de pondérer l’influence des différents éléments d’apprentissage dans le choix de la classe d’un nouvel élément. Wettschereck propose un état de l’art de ces techniques [Wettschereck1997].

Une fois la méthode de calcul de distance choisie, il est nécessaire de déterminer le nombre k d’éléments d’apprentissage pris en compte pour la classification d’une nouvelle donnée.

4.2.1.3 Choix de k

Le choix de l’hyperparamètre k est également crucial pour le bon fonctionnement de l’algorithme. En effet, une trop petite valeur risque d’entraîner du *sur-apprentissage*, c’est-à-dire que l’algorithme ne sera pas capable de bien généraliser sur des données tests, et sera très sensible à des données aberrantes dans l’ensemble d’apprentissage. Ainsi, plus la valeur de k est grande,

plus le risque de sur-apprentissage et l'effet de bruit d'apprentissage sont réduits, mais moins la séparation entre les classes est nette, comme illustré en Figure 4.3.

Dans le cas extrême de $k = 1$, l'algorithme classe les nouveaux éléments en fonction du voisin le plus proche, et est appelé l'*algorithme du plus proche voisin*. Dans le cas opposé, si $k = N$ avec N le nombre de données d'apprentissage, l'algorithme classe tout nouvel élément selon la classe majoritaire dans l'ensemble d'apprentissage.

Le choix du meilleur k est réalisé par des méthodes d'optimisation d'hyperparamètres. Parmi ces techniques, la validation croisée est souvent utilisée. Elle consiste à séparer l'ensemble des données d'apprentissage en un sous-ensemble d'entraînement et un de test. Cela permet de chercher une valeur optimale pour k en minimisant l'erreur de classification sur le sous-ensemble de test. Afin d'optimiser au maximum ce choix, plusieurs itérations de validation croisée sont souvent nécessaire en faisant varier la composition du sous-ensemble de test. La valeur finale de k est alors choisie en combinant les résultats des différentes itérations.

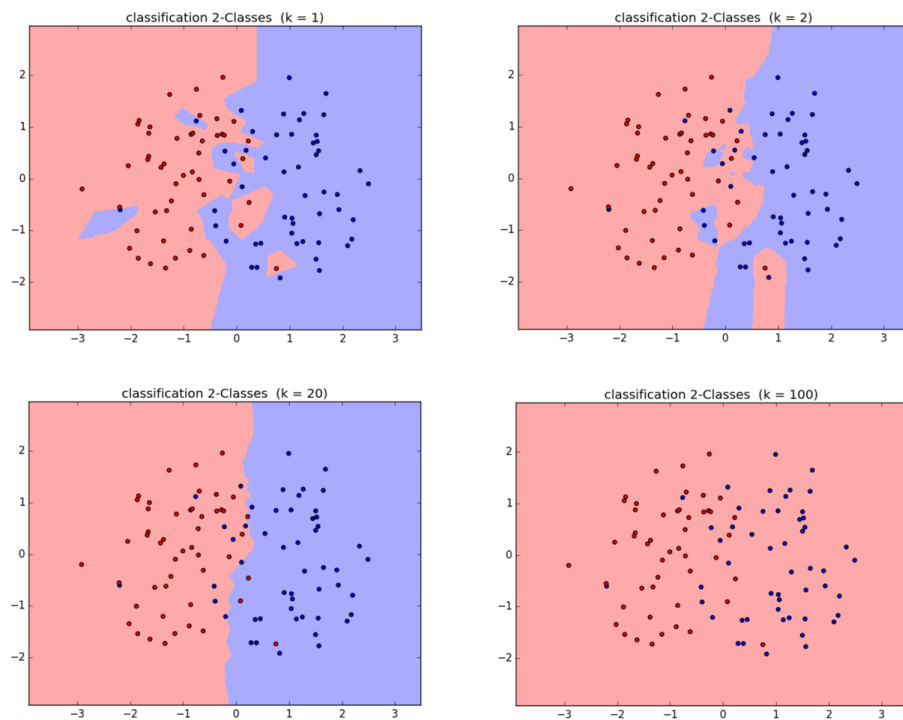


Figure 4.3 – Effet du choix de k pour un problème de classification à 2 classes - Haut/Gauche : $k = 1$, Haut/Droite : $k = 2$, Bas/Gauche : $k = 20$, Bas/Droite : $k = 100$ - Exemple issu de [Kiran2016]

4.2.1.4 Forces et limites

Quelques forces et limites de la méthode des k -plus-proches voisins sont présentées, sur la base des concepts exposés précédemment.

Parmi les atouts majeurs de ces techniques, on retrouve :

- ⇒ L'intuitivité de son fonctionnement qui permet une implémentation rapide et une simplicité d'utilisation.
- ⇒ Sa robustesse aux données aberrantes si le choix de k est assez grand.
- ⇒ Sa capacité naturelle à traiter des problèmes multi-classes.

Parmi les difficultés majeures de cette technique, on retrouve :

⇒ Sa lourdeur en terme de temps de calcul et de place occupée en mémoire. En effet, toutes les valeurs de l'ensemble d'apprentissage doivent être stockées en mémoire pour pouvoir réaliser la classification d'un nouvel élément. De plus, le calcul des distances aux plus proches voisins ainsi que l'évaluation de la classe majoritaire est réalisée à chaque fois, ce qui entraîne un temps de calcul d'autant plus grand que les exemples d'apprentissage sont nombreux. C'est un désavantage majeur de cette technique par rapport aux méthodes qui construisent un modèle statistique lors de l'apprentissage. Ces dernières ne nécessitent pas de stockage de l'ensemble des éléments d'apprentissage, et la classification d'un nouvel élément est généralement rapide une fois le modèle entraîné.

Néanmoins, certaines techniques permettent de réduire ce temps de calcul, notamment en réduisant le nombre d'éléments d'apprentissage, afin de ne garder que les plus significatifs et éliminer l'information redondante.

⇒ Sa sensibilité au phénomène de fléau ou malédiction de la dimension (*curse of dimensionality* en anglais). Si la méthode fonctionne bien pour des données à dimensions faibles, elle risque de produire de moins bonnes performances sur des données à plus grandes dimensions. En effet, plus le nombre de dimensions augmente, plus le volume de l'espace augmente, et plus les données se retrouvent isolées et éparses. Les notions de distances perdent alors de leur significativité.

Un moyen de résoudre ce problème est d'utiliser des méthodes de réduction dimensionnelle, comme Analyse en Composantes Principales (ACP), avant d'utiliser l'algorithme sur des données à grandes dimensions.

4.2.2 Réseaux bayésiens naïfs

Les réseaux bayésiens naïfs sont des approches de classification probabilistes basées sur le théorème statistique de Bayes. Ils sont un cas simple de réseaux bayésiens, qui sont eux mêmes des modèles probabilistes graphiques représentant un ensemble de variables aléatoires et les dépendances conditionnelles les reliant entre elles.

4.2.2.1 Théorème de Bayes

Le théorème de Bayes est issu des travaux des mathématiciens Thomas Bayes et Richard Price, publiés à titre posthumum en 1763 [Bayes1763]. Les résultats de ces recherches ont été étendus par Pierre-Simon Laplace en 1774, qui en a ensuite publié sa formulation moderne quelques dizaines d'années plus tard dans sa *Théorie analytique des probabilités* [Laplace1820].

La formulation la plus commune du théorème de Bayes est décrite par l'Equation 4.1. Considérant deux *événements*, cette formule permet de déterminer la *probabilité conditionnelle* de A sachant B, c'est-à-dire que A se produise *sous condition* que B se soit produit.

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)} \quad (4.1)$$

$P(A)$ est la probabilité *a priori* de A, c'est-à-dire "antérieure" à toute information sur B. De même, $P(B)$ est la probabilité *a priori* de B. Ces probabilités sont aussi quelques fois appelées probabilités *marginales* dans la littérature.

$P(A|B)$ est la probabilité de A sachant B, qualifiée "postérieure" au sens qu'elle dépend de l'information sur B.

$P(B|A)$ est la probabilité de B sachant A. Elle est aussi appelée la *fonction de vraisemblance* de A, pour un B connu.

Les statistiques bayésiennes sont basées sur ce théorème et permettent donc de déterminer la probabilité d'un événement à partir de celles d'autres événements déjà évalués. Les probabilités sont interprétées comme un degré de croyance en la réalisation d'un événement s'appuyant sur des distributions de probabilités construites à partir d'observations empiriques.

A partir de ces principes, il est possible de construire des modèles graphiques représentant les variables aléatoires étudiées et les dépendances conditionnelles qui les lient. Ces représentations sont appelées des réseaux bayésiens.

4.2.2.2 Réseaux bayésiens

Les réseaux bayésiens se présentent sous la forme de *graphes orientés acycliques* (*directed acyclic graph* en anglais ou *DGA*). Le qualificatif *orienté* implique qu'ils sont constitués de nœuds et d'arcs fléchés les reliant entre eux. Le terme *acyclique* signifie qu'aucune suite d'arcs consécutifs ne possède d'extrémités identiques, c'est-à-dire qu'aucun cycle ne peut exister dans le modèle. Les nœuds représentent des variables aléatoires alors que les arcs représentent les relations de dépendances statistiques entre ces dernières.

La Figure 4.4 présente un réseau bayésien simple avec les tables représentant les distributions de probabilité indiquant comment la probabilité qu'une variable prenne une certaine valeur dépend des valeurs prises par ses variables précédentes. Dans l'exemple, les variables peuvent prendre des valeurs binaires et la probabilité que chacune d'elle prenne une certaine valeur est donnée en fonction des combinaisons des valeurs prises par les variables parentes. Le réseau caractérise donc chaque événement de manière probabiliste et non déterministe.

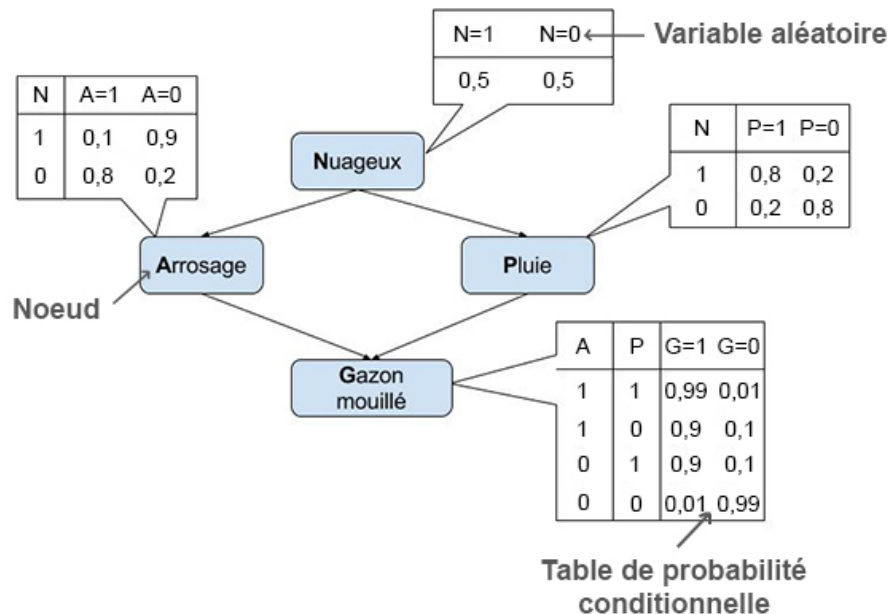


Figure 4.4 – Exemple de réseaux bayésien et des tables de probabilités associées [CNRS2016]

L'utilisation de ce type de réseau permet de calculer les probabilités d'événements non observés, grâce à des données observées. On parle d'*inférence*. De plus les calculs mis en œuvre sont simples car ils ne sont composés que de produits et d'additions.

Ce processus justifie l'intérêt d'utiliser ce type de réseaux comme des classifieurs.

4.2.2.3 Classification par réseaux bayésiens naïfs

Nous décidons de présenter les principes de fonctionnement de la classification à partir de réseaux bayésiens dits *naïfs*. Il s'agit du type le plus simple de classification par réseaux de Bayes et un des plus utilisés dans la littérature.

Après une présentation rapide du modèle bayésien naïf, nous exposerons les processus entrant en jeu dans l'apprentissage d'un tel modèle et sur le principe de classification d'un nouvel élément au travers du modèle entraîné.

Modèle bayésien naïf

Le modèle bayésien naïf est seulement composé de 2 niveaux. Le premier contient un nœud parent C qui correspond aux classes possibles du modèle. Le second niveau contient quant à lui plusieurs nœuds D correspondant aux descripteurs utilisés pour caractériser les éléments à classer. Chaque nœud secondaire n'a pour seul parent que le nœud C . Une représentation de cette structure est présentée en Figure 4.5.

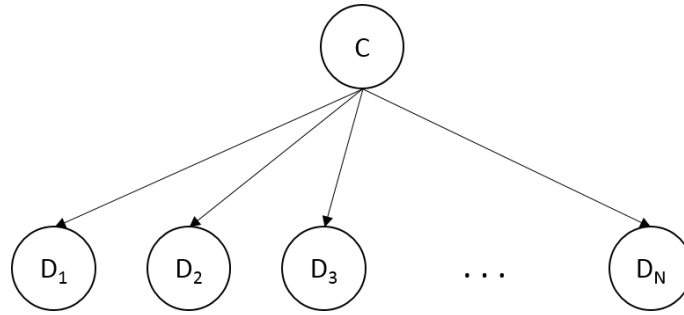


Figure 4.5 – Représentation d'un réseau bayésien naïf

Ces réseaux sont qualifiés de naïfs car ils font l'hypothèse que tous les nœuds fils sont *indépendants* entre eux, pour une valeur de classe donnée. Il s'agit d'une indépendance probabiliste aussi appelée *indépendance conditionnelle*. Elle implique que pour des événements A et B pour un C donné, $P(A|B, C) = P(A|C)$ pour toutes les valeurs de A , B et C , quand $P(C) > 0$.

Si cette hypothèse semble peut vraisemblable en réalité, ces classifieurs se révèlent cependant fonctionnels en pratique et aussi performants que d'autres algorithmes couramment utilisés pour des problèmes donnés. Les raisons de ces performances malgré la violation de l'hypothèse sont explorées par Zhang [Zhang2004].

Si on considère un élément à classer exprimé par un vecteur $\vec{D} = (D_1, D_2, \dots, D_N)$ représentant N descripteurs, sa probabilité d'être assigné à une classe C_k est

$$P(C_k|D_1, D_2, \dots, D_N) = P(C_k|\vec{D}) \quad (4.2)$$

pour toutes k classes possibles.

A l'aide du théorème de Bayes, cette probabilité peut être exprimée selon les probabilités $P(C)$, $P(D_1, D_2, \dots, D_N|C)$ et $P(D_1, D_2, \dots, D_N)$:

$$P(C_k|D_1, D_2, \dots, D_N) = \frac{P(C) \times P(D_1, D_2, \dots, D_N|C)}{P(D_1, D_2, \dots, D_N)} \quad (4.3)$$

En prenant en compte la loi des *probabilités jointes* et de l'hypothèse d'indépendance faite précédemment, cette formule devient :

$$P(C_k|D_1, D_2, \dots, D_N) = \frac{P(C) \times P(D_1|C) \times P(D_2|C) \times \dots \times P(D_N|C)}{P(D_1, D_2, \dots, D_N)} \quad (4.4)$$

$$= \frac{1}{Z} \times P(C) \times \prod_{i=1}^N P(D_i|C) \quad (4.5)$$

avec $Z = P(D_1, D_2, \dots, D_N)$, terme constant car il ne dépend pas de C , mais seulement des variables D_i connues. Il est quelque fois appelé *évidence*.

A partir des principes de ce modèle, il est possible de construire un classifieur dont l'apprentissage et le fonctionnement sont exposés dans la suite de cette section.

Apprentissage et estimation

La construction du classifieur à partir du modèle de probabilité consiste tout d'abord en l'estimation des paramètres du modèle :

- Probabilités *a priori* des classes $P(C)$: elles peuvent être calculées sur l'hypothèse de classes équiprobables, c'est-à-dire $P(C) = \frac{1}{k}$, ou sur la base de l'ensemble des données d'apprentissages, c'est-à-dire $P(C) = \frac{\text{nombre d'exemples dans } C}{\text{nombre d'exemples total}}$;
- Paramètres de la loi de probabilité présumée pour les données : de manière générale la loi normale est très utilisée, et son espérance et sa variance sont estimées pour chaque descripteur D relativement à chaque classe.

En réalité, l'estimation de ces paramètres permet seulement de caractériser le modèle de probabilité. Afin de disposer d'un véritable classifieur, il est nécessaire d'ajouter une *règle de décision*. Celle-ci permet d'assigner une classe à un nouvel élément à partir des probabilités calculées.

La règle du *maximum a posteriori* est souvent utilisée. Elle stipule que l'élément appartient sûrement à la classe dont la probabilité associée est la plus grande.

Le classifieur résultant est défini par l'équation suivante :

$$Y = \operatorname{argmax}_k P(C_k) \prod_{i=1}^N P(D_i|C_k) \quad (4.6)$$

Afin de mieux comprendre le fonctionnement en pratique de ce type de classifieur, nous invitons le lecteur à visiter la [page wikipedia dédiée à la classification naïve bayésienne](#), qui propose un exemple simple et parlant à propos d'une classification d'individu par sexes, en fonction de caractéristiques mesurées (taille, poids, pointure).

4.2.2.4 Forces et limites

Quelques forces et limites de la méthode des réseaux bayésiens naïfs sont présentées, sur la base des concepts exposés précédemment.

Parmi les atouts majeurs de ces techniques, on retrouve :

- ⇒ Sa facilité de mise en œuvre pour la construction du modèle et l'estimation des paramètres.
- ⇒ Sa rapidité d'exécution même sur de très grands ensembles de données.
- ⇒ Sa capacité à estimer les paramètres même avec peu de données d'apprentissage.
- ⇒ Sa capacité à traiter des problèmes de classification multiclassés.

- ⇒ Sa qualité de classifieur probabiliste, qui permet d'obtenir non seulement une classe pour un nouvel élément, mais également un degré de certitude.
- ⇒ Sa robustesse au fléau de la dimension de part l'hypothèse d'indépendance entre les descripteurs.

Parmi les difficultés majeures de cette technique, on retrouve :

- ⇒ Son hypothèse d'indépendance entre les descripteurs, qui le sont rarement en réalité. Cela peut avoir pour conséquence de mener à des résultats erronés, d'où sont appelées de classification "naïve". Il est donc impossible d'apprendre des interactions entre les descripteurs. Néanmoins, il a été montré que ces classifieurs peuvent se révéler optimaux pour certains problèmes donnés malgré la violation de l'hypothèse.
- ⇒ Il s'agit d'un classifieur linéaire, il ne pourra donc pas modéliser des relations trop compliquées entre les données.

4.2.3 Machines à Vecteurs de Support - Support Vector Machine (SVM)

Introduits en 1995 par Vapnik, les SVM font partis de la famille des algorithmes d'apprentissage supervisé [Vapnik1995]. Différents livres et articles ont depuis développé les principes et présenté de nombreuses applications de cette méthode comme [Bennett2000] [Cristianini2000], ou plus récemment [Steinwart2008].

Les SVM permettent dans leur forme primaire de traiter des problèmes de classification binaire, c'est à dire ne comportant que 2 classes. Certaines évolutions de l'algorithme initial sont exposées un peu plus loin pour le traitement de problèmes multi-classes, après la description des principes de fonctionnement généraux.

4.2.3.1 Cas d'une séparation linéaire

Le principe de fonctionnement des SVM repose sur le choix d'un **hyperplan** séparant les données dans un espace multidimensionnel. Considérons tout d'abord le cas où il existe une séparation linéaire entre les données, qui correspond au problème initial posé par Vapnik.

Soit $(\vec{x}_1, y_1), (\vec{x}_2, y_2), \dots, (\vec{x}_N, y_N)$ un ensemble de données d'entraînement de taille N , composé de vecteurs d'entrée $\vec{x}_i$ de dimension p et de leur classe de sortie $y_i \in \{1, -1\}$. Tout hyperplan qui sépare le groupe de point $\vec{x}_i$ pour lesquels $y_i = 1$ du groupe de points $\vec{x}_i$ pour lesquels $y_i = -1$ est appelé *hyperplan séparateur* et répond à l'équation (4.7).

$$\vec{w} \cdot \vec{x} + b = 0 \quad (4.7)$$

avec $\vec{w}$ vecteur normal à l'hyperplan, composé des poids w_i . Le paramètre $\frac{b}{\|\vec{w}\|}$ caractérise l'écart de l'hyperplan par rapport à l'origine, suivant $\vec{w}$, comme illustré en 2 dimensions sur la Figure 4.6.

Dans le cas où les données sont linéairement séparables, il existe néanmoins une multitude d'hyperplans permettant de réaliser la séparation, comme représenté en Figure 4.6a. L'hyperplan optimal est alors obtenu en maximisant la **marge** qui sépare le plan des échantillons de données.

La marge correspond à la distance entre deux hyperplans parallèles passant par les échantillons de données les plus proches de chaque groupe (Figure 4.6). Ces points particuliers sont appelées **vecteurs supports** et donnent leur nom à la méthode. Les deux hyperplans passant par ces points sont décrits par les équations (4.8) et (4.9). L'hyperplan optimal se situe au milieu de ces deux hyperplans.

$$\vec{w} \cdot \vec{x} + b = 1 \quad (4.8)$$

et

$$\vec{w} \cdot \vec{x} + b = -1 \quad (4.9)$$

Tous les points au delà de l'hyperplan d'équation (4.8) font partis de la classe 1 alors que tous les points au delà de l'hyperplan d'équation (4.9) font partis de la classe -1.

La distance entre ces deux hyperplans est alors de $\frac{2}{\|\vec{w}\|}$. Pour maximiser la distance qui les sépare, donc maximiser la séparation entre les données, il faut donc minimiser $\|\vec{w}\|$. La distance utilisée est la distance d'un point à un plan qui renvoie à la plus petite distance entre le point considéré et un point du plan.

De plus, une autre contrainte est ajoutée pour garantir que chaque vecteur de donnée appartient bien au bon côté de l'hyperplan. Elle peut être décrite par l'équation (4.10).

$$y_i(\vec{w} \cdot \vec{x} + b) \geq 1, \text{ avec } 1 \leq i \leq N \text{ et } y_i \in \{1, -1\} \quad (4.10)$$

La détermination de l'hyperplan optimal se résume alors en un problème d'optimisation cherchant à minimiser $\|\vec{w}\|$ en respectant la condition (4.10).

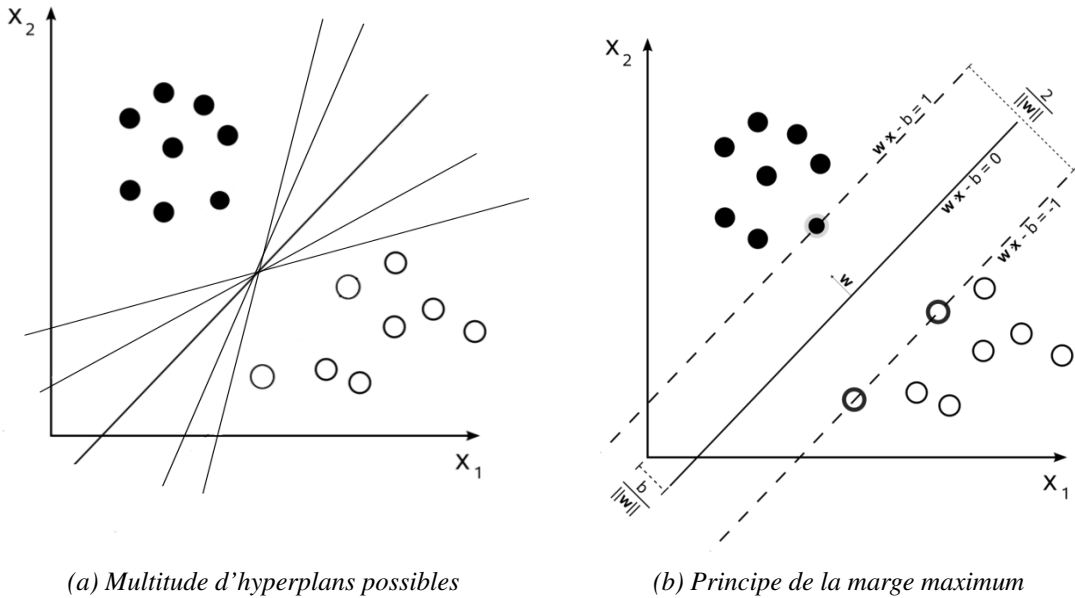


Figure 4.6 – Illustrations des principes d'hyperplans et marges maximales

4.2.3.2 Cas d'une séparation non-linéaire

L'algorithme initial introduit précédemment présente une forte limitation dans le fait qu'il ne fonctionne seulement dans les cas où les données sont linéairement séparables. Plusieurs évolutions ont été cependant proposées pour pouvoir couvrir les cas de séparations non-linéaires.

Changement d'espace - Astuce du noyau ou "kernel trick"

La première solution réside dans la construction d'un classifieur non-linéaire en changeant d'espace vectoriel. Succinctement, cette méthode est basée sur l'idée qu'un problème non séparable linéairement dans l'espace vectoriel des données initiales l'est peut être dans un espace vectoriel transformé, appelé **espace de redescription**. L'illustration en Figure 4.7 permet de comprendre ce principe sur un exemple simple. Des données non séparables linéairement dans l'espace des

coordonnées cartésiennes, le deviennent en passant dans l'espace des coordonnées polaires. Dans ce cas, la dimension de l'espace de départ est la même que celle de l'espace d'arrivée. Dans le cas des SVMs, le problème peut être reconsidéré dans un espace de **dimension supérieure**, voir infinie.

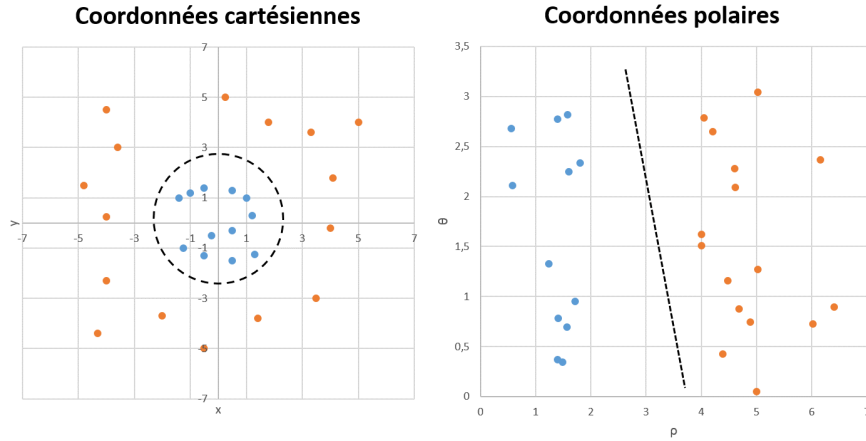


Figure 4.7 – Exemple de changement d'espace en passant de coordonnées cartésiennes à des coordonnées polaires

Si on appelle Φ la transformation appliquée pour passer de l'espace de départ à l'espace transformé de dimension supérieure, appelé parfois espace- Φ , alors le vecteur :

$$(\vec{x}_1, y_1), (\vec{x}_2, y_2), \dots, (\vec{x}_N, y_N)$$

devient :

$$(\Phi(\vec{x}_1), y_1), (\Phi(\vec{x}_2), y_2), \dots, (\Phi(\vec{x}_N), y_N)$$

dans le nouvel espace.

Le problème d'optimisation posé précédemment peut alors être appliqué pour estimer l'hyperplan optimal dans le nouvel espace, en prenant les $\Phi(\vec{x}_i)$ au lieu des $\vec{x}_i$.

Néanmoins, la résolution de ce problème se heurte alors à ce qui est appelé en statistiques la *malédiction de la dimensionnalité*, et qui rapporte que la difficulté d'un problème d'estimation augmente dramatiquement avec la dimension de l'espace. En effet, le nombre de calculs nécessaires pour construire les nouvelles données dans l'espace- Φ , et pour leur appliquer tous les produits vectoriels nécessaires à résoudre l'optimisation devient vite insurmontable.

Pour surmonter cet obstacle, Boser, Guyon et Vapnik ont eu l'idée d'appliquer le principe dit de l'*astuce du noyau*, ou *kernel trick* en anglais, [Boser1992]. Ce principe est par ailleurs clairement expliqué par Muller dans un état de l'art de différents algorithmes utilisant cette astuce [Muller2001].

Le fondement de cette astuce réside dans le fait qu'il est directement possible de calculer les produits scalaires dans l'espace- Φ , sans jamais avoir à calculer explicitement les $\Phi(\vec{x}_i)$. Le produit scalaire des vecteurs dans l'espace- Φ est alors calculé via une fonction particulière, appelée **fonction noyau** et généralement notée K . Pour pouvoir être utilisée dans ce contexte, une fonction noyau doit répondre au *critère de Mercer* qui stipule que $K(a, b)$ doit être continue, symétrique et définie positive [J Mercer1909].

Parmi les noyaux les plus utilisées, on retrouve les noyaux à fonction polynomiale, à base radiale gaussienne (RBF pour *Radial Basis Function*), sigmoïde ou encore quadratique inverse [Muller2001].

Marge souple ou Soft Margin

Dans le cas où il reste toujours des erreurs dans la séparation des données d'entraînement, c'est-à-dire des points se trouvant du mauvais côté de l'hyperplan, Cortes et Vapnik ont également introduit le concept de marge souple (ou *soft margin* en anglais) [Cortes1995]. Cette technique permet de mieux tolérer les mauvais classements, telles que des valeurs aberrantes ou des erreurs d'étiquetage, dans la recherche de l'hyperplan optimal.

De manière simple, de nouvelles variables ξ_i sont introduites pour permettre à l'algorithme de faire un petit nombre d'erreurs dans la détermination de l'hyperplan optimal. Ces variables sont appelées variables ressort ou *slack variables*, et caractérisent les distances séparant les points mal classés de l'hyperplan (Figure 4.8). Plus concrètement, si ξ_i n'est pas nul, cela veut dire que x_i ne correspond pas avec la marge fixée, à un coût proportionnel à ξ_i . La somme $\sum_{k=1}^p \xi_i$ caractérise alors le nombre d'erreurs d'apprentissage.

Le problème d'optimisation précédent devient alors :

$$\text{Minimiser } \frac{1}{2}(\|w\|^2) + C \sum_{k=1}^p \xi_i \text{ avec } C \geq 0 \quad (4.11)$$

$$y_i(\vec{w} \cdot \vec{x} + b) \geq 1 - \xi_i, \text{ avec } 1 \leq i \leq N, y_i \in \{1, -1\} \text{ et } \xi_i \geq 0 \quad (4.12)$$

La constante C est le terme de *régularisation* qui permet de contrôler le compromis entre le nombre d'erreurs d'apprentissage et la largeur de la marge. Il s'agit alors de déterminer sa valeur pour minimiser le nombre d'erreurs tout en gardant une marge maximale.

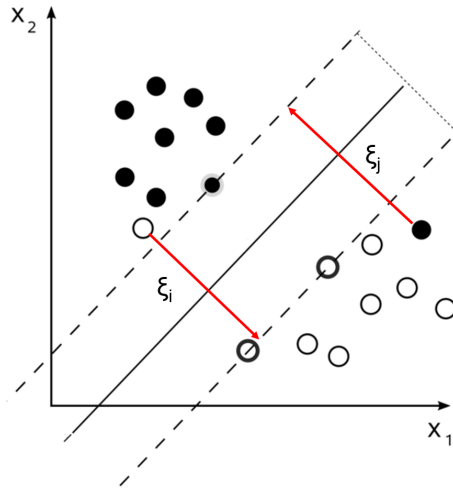


Figure 4.8 – Illustration des variables ressort ξ_i

4.2.3.3 Cas d'une classification multi-classes

Bien que la méthode initiale ne concernait que des problèmes de classification binaires, des techniques ont également été proposées pour étendre l'utilisation des SVMs à des problèmes multi-classes (nombre de classes supérieur à 2). De nombreux schémas ont été développés et sont décrits dans un état de l'art récent de ces techniques proposé par Wand et Xue [Wang2014]. Pour donner un aperçu du fonctionnement possible de tels algorithmes, nous présentons succinctement deux approches parmi les plus classiquement utilisées : l'approche un contre tous ou *one-versus-all* ou *one-versus-rest* et l'approche un contre un ou *one-versus-one*.

L'approche *one-versus-rest* construit k classifieurs binaires pour un problème comportant k classes.

Chaque classifieur m prend les éléments d'apprentissage de la classe m comme label 1 et les éléments des $k - 1$ classes restantes comme label -1. Pendant la phase de test, la classe d'un élément est donnée par le classifieur proposant la valeur de sortie la plus élevée.

L'approche *one-versus-one* construit $k(k - 1)/2$ classifieurs binaires, afin de confronter une à une chacune des k classes. Un système de vote est alors mis en place lors du test d'un élément. Chaque classifieur vote pour une classe, et l'élément est attribué à celle qui remporte le plus de voix au final. Le résultat peut néanmoins être ambigu si il n'existe pas de vote majoritaire.

4.2.3.4 Forces et limites

Pour finir, quelques forces et limites des SVMs sont présentées, sur la base des concepts présentés précédemment.

Parmi les atouts majeurs de ces techniques, on retrouve :

- ⇒ Leur capacité à traiter des problèmes présentant des données à grandes dimensionnalités.
- ⇒ Leur capacité à traiter des problèmes non-linéaires grâce à l'introduction des fonctions noyau.
- ⇒ Leur robustesse aux points aberrants grâce à la stratégie de la marge souple.
- ⇒ Peuvent être adaptés à des problèmes multi-classes.

Parmi les difficultés majeures de ces techniques, on retrouve :

- ⇒ La difficulté à choisir la bonne fonction noyau pour un problème donné.
- ⇒ Complexité et longueur de l'entraînement pour des grands jeux d'entraînement.
- ⇒ Difficulté d'interprétation de la prise de décision à travers le modèle final.

4.2.4 Arbres décisionnels et Forêts d'arbres décisionnels

Les arbres décisionnels font partis de la catégorie des algorithmes d'apprentissage supervisé, et ont été décrits par de nombreux auteurs. L'état de l'art proposé par Murphy en 1998 permet notamment d'avoir un bon aperçu des travaux menés autour de ces techniques [Murthy1998]. Après en avoir décrit les principes, nous aborderons également certaines extensions de cette méthode permettant d'en améliorer les performances : l'ensachage d'arbres (ou *bagging* en anglais), le boosting d'arbres et la classification par forêt d'arbres décisionnels (ou *random forest*) proposée par Breiman [Breiman2001].

4.2.4.1 Classification par arbre de décision

Arbres et règles de décision

Un arbre décisionnel est un classifieur représenté sous forme d'arbre qui permet d'attribuer une classe à un individu en fonction de tests successifs menés sur certains de ces attributs. Il peut être aussi bien utilisé avec des attributs quantitatifs que qualitatifs.

Les Figures 4.9a et 4.9b illustrent des cas simples pour appréhender facilement le principe de la technique.

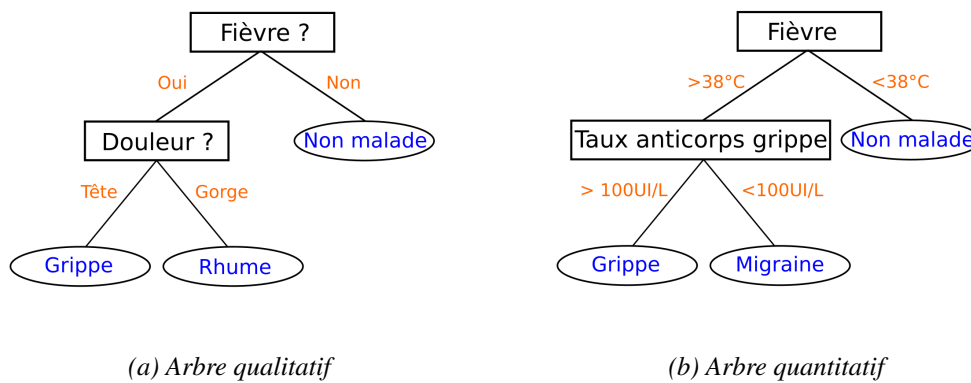


Figure 4.9 – Exemples d'arbres décisionnels simples

On retrouve 3 types d'éléments dans un arbre de décision :

- Les **nœuds** de l'arbre testent les attributs et sont représentés par les cases noires sur les figures.
- Les **feuilles** de l'arbre correspondent aux classes et sont représentées par des ellipses bleues sur les figures.
- Les **branches** de l'arbre portent les valeurs possibles des attributs (en orange sur les figures) testés. Chaque branche guide la décision vers un nouveau test ou vers une classe. Les valeurs peuvent être aussi bien qualitative que quantitatives.

Les arbres de décision étant des représentations graphiques, ils sont souvent traduits en **règles de décision**. L'ordre dans lequel les attributs sont testés est imposé et les règles sont mutuellement exclusives. Chaque règle suit un chemin dans l'arbre en partant du premier nœud, jusqu'à une des feuilles.

Dans le cas des exemples 4.9a et 4.9b, les règles sont traduites ainsi :

- Exemple 4.9a (qualitatif) :

SI Fièvre = Oui **ET** Douleur = Tête **ALORS** Grippe
SI Fièvre = Oui **ET** Douleur = Gorge **ALORS** Rhum
SI Fièvre = Non **ALORS** Nonmalade

- Exemple 4.9b (quantitatif) :

SI Fièvre > 38°C **ET** Tauxanticorpsgrippe > 100UI/L **ALORS** Grippe
SI Fièvre > 38°C **ET** Tauxanticorpsgrippe < 100UI/L **ALORS** Migraine
SI Fièvre < 38°C **ALORS** Nonmalade

A partir d'un ensemble de données d'apprentissage, il est possible d'apprendre un arbre de décision et ses règles. Les principes de cet apprentissage sont introduit dans le prochain paragraphe.

Principes de l'apprentissage d'un arbre de décision

On se place pour la suite dans le cas d'une classification quantitative afin de présenter le principe d'apprentissage des arbres de décision.

De manière générale, l'apprentissage d'un arbre décisionnel repose sur la division successive de

l'ensemble des données d'apprentissage, grâce à des tests sur les attributs, jusqu'à ce que les sous-ensembles contiennent le moins possible d'éléments de classes différentes.

L'ensemble d'apprentissage est divisé tel que la **pureté** des sous-ensembles créés est augmentée à chaque itération. La pureté d'un ensemble est définie par la domination d'une classe parmi les données qui en font parti. Un ensemble parfaitement pur ne contiendrait alors que des données appartenant à une seule et même classe.

En pratique, il reste généralement des erreurs de classification dans l'apprentissage et l'arbre présente une certaine **erreur d'apprentissage** qui caractérise le niveau de pureté de ses différents sous-ensembles. Il s'agit alors de calculer un arbre en minimisant l'erreur d'apprentissage.

La définition d'un arbre décisionnel comprend 3 étapes fondamentales : les tests de séparation, l'arrêt de la croissance et l'attribution des classes aux feuilles. Différents algorithmes ont été développés et diffère dans leur manière d'aborder ces trois étapes. Parmi les plus utilisés, on peut citer l'algorithme CART ("*Classification and Regression Trees*") [Breiman1984] et le C4.5 [Quinlan1993].

Tests de séparation

A partir de l'ensemble des données de départ, des règles de séparation sont définies pour définir des sous-ensembles.

Pour des données d'apprentissage notées $(\vec{x}, y)$, où $\vec{x}$ est un vecteur d'attributs et y la sortie étiquetée correspondante, faisant parties d'un ensemble S , la règle usuelle de séparation binaire en deux sous-ensembles L et $R \in S$ est décrite par 4.13.

$$L = \{(\vec{x}, y) \in S \mid x_d \leq t\} \quad \text{et} \quad R = \{(\vec{x}, y) \in S \mid x_d > t\} \quad (4.13)$$

où x_d est la d -ième composante de $\vec{x}$ et t un nombre réel.

Le processus d'apprentissage consiste alors en l'enchaînement de plusieurs règles binaires, comme illustré en Figure 4.10 pour un problème simple en 2 dimensions. Il est intéressant de noter que même si les règles singulières définissent une séparation binaire, l'ensemble des données peut être séparer en plus de 2 classes grâce à des associations de plusieurs règles.

A chaque itération, dans l'ensemble des tests possibles, la règle de séparation optimale des données d'entraînement de l'espace S en deux sous-ensembles L et R est définie telle qu'elle maximise la diminution de l'erreur d'apprentissage. La variation de cette erreur peut être décrite par l'équation 4.14 .

$$\Delta i(S, L, R) = i(S) - \frac{|L|}{|S|} i(L) - \frac{|R|}{|S|} i(R) \quad (4.14)$$

où $|S|$ est le nombre d'éléments dans S et $i(S)$ correspond à l'erreur d'apprentissage de S , qui caractérise la fraction d'éléments de S dont la classe est différente de la classe la plus fréquente dans S . Elle est parfois aussi nommée l'*impureté* ou l'*homogénéité* de S .

Cette fonction $i(S)$ peut être calculée de différentes manières en fonction des algorithmes. Par exemple, l'algorithme CART utilise l'**indice de pureté de Gini** ou de **diversité de Gini**, décrit en équation 4.15, alors que l'algorithme C4.5 utilise le **gain d'information**, décrit en équation 4.16.

$$i(S) = \sum_{y \in Y} f_{y|S} (1 - f_{y|S}) = 1 - \sum_{y \in Y} f_{y|S}^2 \quad (4.15)$$

où Y est l'ensemble des classes y et $f_{y|S}$ désigne la fraction des éléments de l'ensemble S qui appartiennent à la classe y . Cet indice atteint son minimum 0 quand tous les éléments de l'ensemble appartiennent à la même classe.

$$i(S) = - \sum_{y \in Y} f_{y|S} \log_2(f_{y|S}) \quad (4.16)$$

Le gain d'information est en fait une mesure de différence d'entropie d'information entre l'état de l'ensemble avant et après séparation. Plus le gain est important, plus la séparation permet de réduire l'erreur d'apprentissage entre l'ensemble parent et les sous-ensembles enfants.

Dans tous les cas, le calcul de la meilleure séparation se fait par itération à travers les composantes $d \in \{1, 2, \dots, D\}$ de x en calculant toutes les valeurs possibles du seuil t de l'équation 4.13.

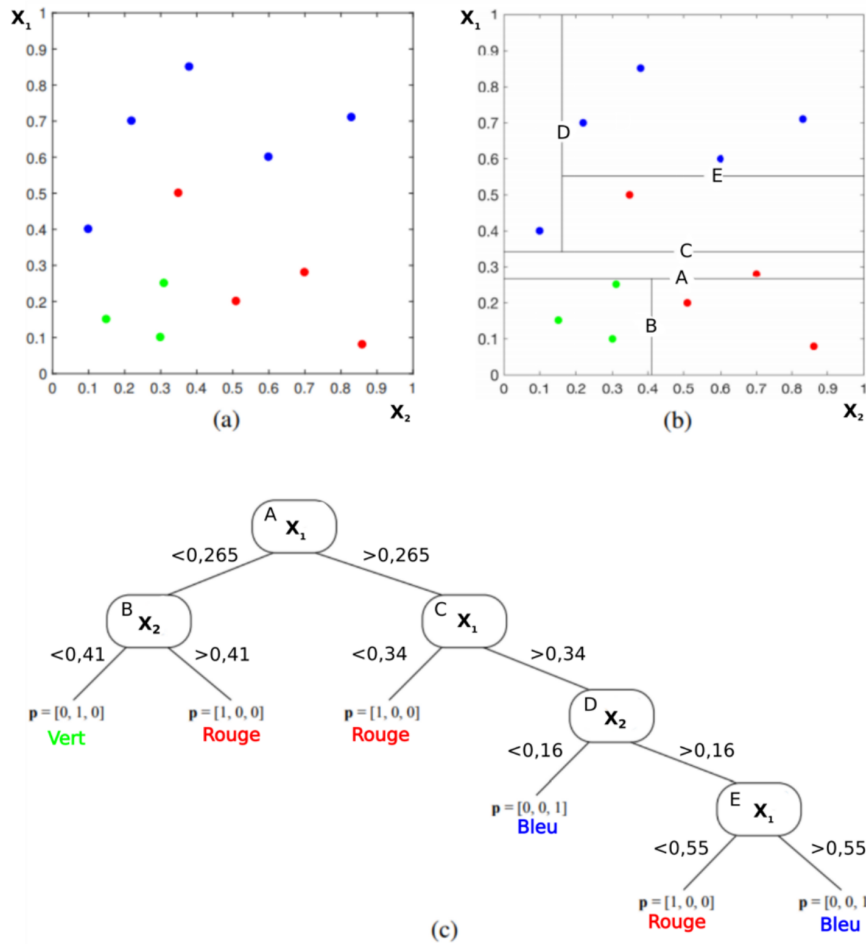


Figure 4.10 – Principe d'apprentissage d'un arbre de classification via un lot d'entraînement divisé en 3 classes (Figure adaptée de [Tomas2017]). Les feuilles peuvent être labélisées soit directement avec la classe (rouge, vert ou bleu), soit avec une distribution des probabilités de chaque classe ($p = [p(\text{rouge}), p(\text{vert}), p(\text{bleu})]$). Dans ce cas, toutes les feuilles contiennent une homogénéité maximum car tous les labels présents dans chacune d'elles appartiennent à une même classe.

Critère d'arrêt

Il est important que la croissance de l'arbre soit limitée afin d'éviter le phénomène de sur-apprentissage. En effet, si l'arbre devient trop grand et trop compliqué, il risque de *trop bien* correspondre aux données d'apprentissage, et ainsi perdre sa capacité à généraliser sur un jeu de données de validation. Il en résulterait un pouvoir de prédiction faible.

D'un autre côté, il ne faut pas non plus couper sa croissance trop tôt, au risque de garder une erreur d'entraînement trop haute, et donc un modèle peu performant dans ce cas également.

Une première solution intuitive serait de stopper la croissance lorsque la diminution d'erreur d'entraînement passe en dessous d'un seuil. Or cela peut se révéler dangereux car il existe des cas dans lesquels une séparation qui ne diminue pas le gain à un moment peut conduire à une diminution plus tard, en étant combinée à une autre séparation, comme l'explique [Tomasi2017].

En réalité, il n'existe pas vraiment de critère d'arrêt optimal. De manière générale, le choix peut être fait de stopper l'algorithme lorsque l'homogénéité d'un sous-ensemble atteint une certaine valeur, ou quand l'arbre atteint un certain nombre de niveau, ou encore lorsqu'une nouvelle séparation mène à un sous-ensemble trop petit.

Ces critères sont généralement fixés de manière empirique ou arbitraire, en fonction de l'application.

Attribution des classes

Généralement, l'attribution des classes s'effectue sur la base de la classe majoritaire présente dans chaque feuille de l'arbre final. La classe majoritaire est alors choisie comme label pour la feuille.

Dans certains cas, il est possible d'exprimer chaque feuille avec sa distribution de probabilité d'apparition de chaque classe, comme sur la Figure 4.10. La probabilité $p(y|S)$ de chaque classe y dans un ensemble S correspond à la fraction d'éléments de classe y dans cet ensemble :

$$p(y|S) = \frac{N_y}{|S|}$$

où N_y est le nombre d'éléments de classe y dans S .

Élagage

Étant donné qu'il est compliqué de trouver un critère d'arrêt optimal pour la méthode présentée jusqu'ici, la plupart de techniques utilisées introduisent une phase d'**élagage** à la suite de la phase de croissance de l'arbre.

Cette méthode consiste à diviser l'ensemble d'entraînement avant l'apprentissage de l'arbre.

Une partie des données est utilisée pour faire grandir l'arbre selon la méthode précédemment présentée, en tentant de maximiser l'homogénéité des feuilles, au risque qu'un phénomène de sur-apprentissage apparaisse, faisant ainsi augmenter l'*erreur de généralisation*.

La seconde partie des données d'apprentissage est ensuite utilisée pour l'élagage. Si différents critères existent pour choisir quelles branches élaguer, le principe reste le même. Différents sous-arbres sont testés avec la seconde partie des données d'apprentissage, et celui qui permet d'augmenter le plus l'erreur de généralisation est retenu.

Il est intéressant de noter que la partie des données utilisée pour la phase d'élagage est appelée échantillon d'élagage (*pruning set* en anglais) pour se démarquer de l'échantillon de validation, utilisé par la suite pour valider les performances globales du modèle.

4.2.4.2 Forces et limites

Quelques forces et limites des arbres de décisions sont présentées, sur la base des concepts présentés précédemment.

Parmi les atouts majeurs de cette technique, on retrouve :

- ⇒ Sa simplicité d'interprétation. Il est en effet facile de remonter les différentes étapes successives pour comprendre comment l'arbre est arrivé à la classe choisie.
- ⇒ Son bon fonctionnement aussi bien avec des données quantitatives que qualitatives.
- ⇒ Sa capacité à traiter des problèmes multi-classes.
- ⇒ Sa capacité à fonctionner avec des données d'amplitudes différentes. Il n'est pas nécessaire de normaliser les entrées par exemple.
- ⇒ Sa capacité à pouvoir traiter des grands ensembles de données avec des ressources de calcul standards.

Parmi les difficultés majeures de cette technique, on retrouve :

- ⇒ Son apprentissage optimal est un problème dit *NP-complet*, ce qui induit que son temps d'exécution augmente de manière exponentielle avec la dimension des données d'entrées. L'algorithme est vite inexploitable même pour une dimension modérée. Une solution est donc en réalité trouvée grâce à des *heuristiques*, c'est-à-dire des méthodes de calcul présentant un résultat rapide, mais pas optimal dans l'absolu.
- ⇒ Son interprétation dans le cas d'arbres très grands et très compliqués. En effet, si il existent trop de conditions pour arriver à une classe, l'interprétation du chemin suivi peut se révéler inexploitable.
- ⇒ Sa sensibilité aux données aberrantes.
- ⇒ Le choix du critère d'arrêt et de la stratégie d'élagage les plus efficaces est délicate.

4.2.4.3 Quelques évolutions des algorithmes d'arbres décisionnels

Afin d'améliorer les performances des arbres décisionnels, différents algorithmes ont été développés sur la base de cette technique. Nous décidons d'en présenter quelque uns parmi les plus utilisés : l'ensachage d'arbres ou *tree bagging*, les forêts d'arbres décisionnels de Breiman et le boosting d'arbres.

Ensachage ou *bagging*

Cette méthode a été appliquée pour la première fois aux arbres décisionnels par Breiman en 1996 [Breiman1996]. Elle consiste à entraîner plusieurs arbres sur des sous-ensembles aléatoires des données d'apprentissage, tirés indépendamment, uniformément et avec remplacement dans l'ensemble initial.

Un grand nombre d'arbres simples sont ainsi générés (de quelques dizaines jusqu'à quelques milliers) et un système de vote mène à la prédiction finale pour un élément. La classe choisie y pour un élément d'entrée x est alors la classe majoritaire après le vote de tous les arbres.

Cette technique permet de corriger le problème de sensibilité aux données aberrantes qui apparaît avec l'utilisation d'un seul arbre décisionnel, en *lissant* la prédiction sur les décisions de plusieurs arbres. Néanmoins, pour que cette technique soit efficace, il est nécessaire que les arbres générés soient les plus dé-corrélés possibles entre-eux. C'est ce qui est réalisé en entraînant les différents arbres sur des lots aléatoires de données d'apprentissage.

Le nombre d'arbres utilisé est souvent déterminé en appliquant des techniques de *validation croisée* et notamment l'erreur *out-of-bag*. Celle-ci correspond à l'erreur de généralisation pour chaque donnée x d'entraînement, calculée seulement en utilisant les arbres n'ayant pas x dans leur sous-ensemble d'entraînement.

Forêt d'arbres décisionnels ou *random forest*

Cette méthode a été présentée par Breiman en 2001 [Breiman2001] et est fondée sur le même principe que l'ensachage d'arbres. Breiman y ajoute une étape supplémentaire pour augmenter la dé-corrélation entre les arbres du modèle. En effet, dans le cas où un ou plusieurs attributs des éléments d'apprentissage influent beaucoup sur la réponse du modèle, cet attribut risque d'être sélectionné dans la plupart des arbres, les rendant fortement corrélés entre eux.

Afin d'éviter ce problème, Breiman introduit une étape de sélection d'un lot d'attributs aléatoires parmi tous les attributs initiaux, avant chaque étape de séparation de l'arbre. Chaque séparation de chaque arbre est donc apprise sur des lots d'attributs différents, ce qui permet de réduire encore plus le risque de corrélation.

A titre informatif, le nombre d'attributs choisi dans les lots est typiquement fixé à $\sqrt{p}$, p étant le nombre d'attributs des éléments x d'apprentissage, pour les problèmes de classification [Tibshirani2008].

Boosting d'arbres décisionnels

Les techniques de *boosting* peuvent être utilisées en combinaison avec différents types de classifieurs, et ont été notamment popularisés par Freund en 1996 avec l'algorithme *AdaBoost* [Freund1996]. Leur application aux arbres décisionnels a été quant à elle proposée en 1996 par Drucker [Drucker1996].

Le principe du boosting repose sur la combinaison des sorties de plusieurs classifieurs dits *faibles* pour produire une prédiction final *forte*. Dans le cadre de l'algorithme *AdaBoost*, un classifieur faible est défini comme un classifieur dont le taux d'erreur est légèrement meilleur qu'une classification aléatoire. Le but du boosting est alors d'appliquer un algorithme de classification faible de manière itérative sur des lots de données pondérées différemment à chaque itération, résultant en la création d'une séquence de classifieurs faibles. La prédiction finale est ensuite calculée à travers un vote majoritaire pondéré des différents classifieurs faibles de la séquence.

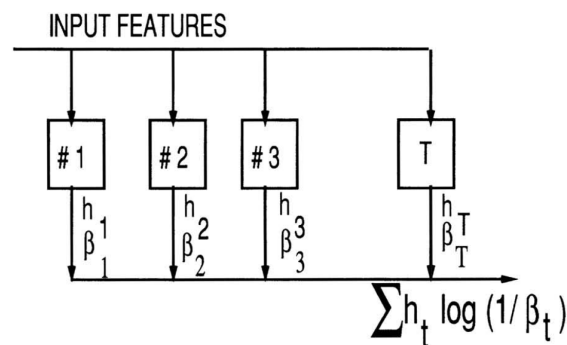


Figure 4.11 – Structure de boosting, issue de [Drucker1996]

Dans sa proposition d'application de l'algorithme AdaBoost aux arbres décisionnels, Drucker présente l'algorithme en Figure 4.11, représentatif de la structure AdaBoost. Chaque boîte contient un arbre de décision de type C4.5 qui joue le rôle de classifieur faible. Le premier arbre est entraîné sur un sous-ensemble d'éléments d'apprentissage tiré aléatoirement avec remise dans l'ensemble complet d'apprentissage. Après entraînement et élagage, l'ensemble complet d'apprentissage est passé au travers du premier arbre. Les éléments mal prédits se voient alors affectés d'un poids qui leur donnera une probabilité plus forte d'être sélectionnés pour l'entraînement du prochain arbre. De cette manière, les nouveaux arbres sont de plus en plus probablement entraînés sur des

éléments difficiles à classer correctement. La sortie globale est ensuite calculée en sommant les sorties h_T de chaque classifieur faible avec un coefficient $\log \frac{1}{\beta_t}$ où β_t est proportionnel à l'erreur d'apprentissage de l'arbre t . Ainsi, si l'arbre t présente une erreur d'apprentissage faible, il va contribuer fortement à la sortie finale car son $\frac{1}{\beta_t}$ va être élevé.

4.2.5 Réseaux de neurones

De nombreuses techniques sont regroupées à l'heure actuelle sous l'appellation de *réseau de neurones*. Elles font références à des algorithmes inspirés du fonctionnement des neurones biologiques. Une description très complète des principes élémentaires qui fondent ces modèles est présentée dans un cours complet et clair de F.Denis et R.Gilleron à l'université de Lille, disponible en ligne [Gilleron2018]. La présentation des différents concepts dans cette section est largement issue et adaptée de ce cours.

Afin de comprendre au mieux ce type de modèle, il est tout d'abord nécessaire de décrire brièvement la structure d'un neurone artificiel.

4.2.5.1 Neurone artificiel ou formel

La Figure 4.12 montre un parallèle entre la structure d'un neurone biologique et celle d'un neurone artificiel. Les synapses sont modélisées par un ensemble d'entrées x pondérées de poids w qui sont ensuite traitées par une combinaison de deux fonctions, celle de combinaison et celle d'activation, qui représentent le traitement par le corps cellulaire du neurone biologique.

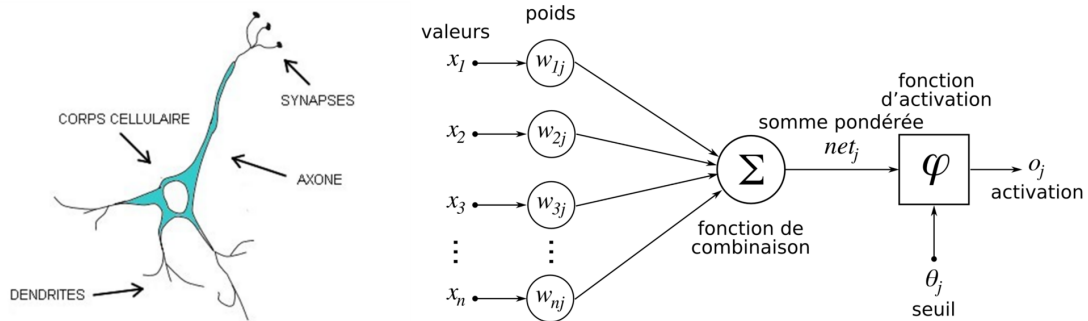


Figure 4.12 – Structure d'un neurone biologique à gauche et artificiel à droite [CC : By Jimmy-neuron - Own work, CC BY-SA 3.0 et Par Chrislb - Made by Chrislb, CC BY-SA 3.0]

La *fonction de combinaison* a, comme son nom l'indique, le rôle de combiner les données d'entrées en un coefficient. Cette combinaison peut être réalisée de différentes manières, comme par une somme pondérée dans le cas le plus classique, que nous utiliserons dans la suite de ces explications. La sortie de cette fonction peut être appelée *portentiel post-synaptique* et est calculée via l'équation 4.17.

$$net = \sum_{i=1}^N w_i x_i \quad (4.17)$$

Cette valeur passe ensuite à travers la *fonction d'activation* qui renvoie un résultat en sortie du neurone. Dans le cas d'un neurone à seuil comme décrit en Figure 4.12, cette fonction est un seuil de paramètre θ qui renvoie un résultat binaire $\{0,1\}$. La valeur o de sortie du neurone est alors calculée via l'équation 4.20.

$$o = \begin{cases} 1 & \text{si } \sum_{i=1}^N w_i x_i > \theta, \\ 0 & \text{sinon} \end{cases} \quad (4.18)$$

Pour simplifier ce modèle, il est d'usage de remplacer ce seuil θ par l'ajout d'une nouvelle entrée $x_0 = 1$, associée à un poids $w_0 = -\theta$, aussi appelé *coefficient de biais*. La fonction de combinaison devient alors :

$$net = \sum_{i=0}^N w_i x_i = \left(\sum_{i=1}^N w_i x_i \right) - w_0 \quad (4.19)$$

avec x_i l'entrée i parmi N du neurone et w_i son poids associé.

La fonction d'activation utilisée est alors l'échelon d'Heaviside (Figure 4.13a), et la sortie o du neurone est alors calculée via l'équation 4.20.

$$\begin{cases} 1 & \text{si } \sum_{i=0}^N w_i x_i = \left(\sum_{i=1}^N w_i x_i \right) - w_0 > 0, \\ 0 & \text{sinon} \end{cases} \quad (4.20)$$

Des fonctions plus compliquées peuvent également être utilisées pour introduire de la non-linéarité dans le modèle. Dans ce cas, la sortie de la fonction d'activation sera un nombre réel dans un intervalle (souvent $[0,1]$) et non plus une sortie binaire. Parmi les fonctions de ce type les plus utilisées, on retrouve la fonction sigmoïde (Figure 4.13b).

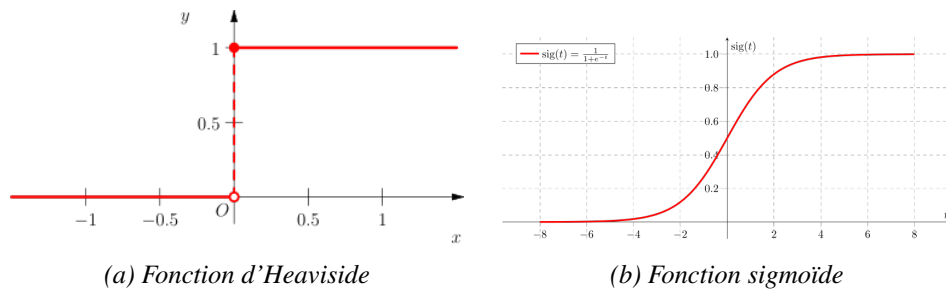


Figure 4.13 – Exemples de fonctions d'activation usuelles

Les fonctions d'activation possèdent un seuil qui définit si le neurone est *actif* ou non. Il est atteint ou dépassé lorsque le potentiel post-synaptique $net \geq 0$, c'est à dire quand la somme pondérée des entrées est égale au coefficient de biais (voir Équation 4.21).

$$net \geq 0 \leftrightarrow \sum_{i=0}^N w_i x_i \geq 0 \leftrightarrow \sum_{i=1}^N w_i x_i - w_0 \geq 0 \leftrightarrow \sum_{i=1}^N w_i x_i \geq w_0 \quad (4.21)$$

Dans le cas d'une fonction de Heaviside, l'équation $\sum_{i=1}^N w_i x_i = w_0$ définit alors un hyperplan dans l'espace des données d'entrée. Selon que l'entrée x appartient à l'un ou l'autre des demi-espaces séparés par l'hyperplan, la fonction d'activation renvoie $f(x)$ égale à 0 ou 1. L'utilisation d'un fonction comme la sigmoïde introduit un peu de non-linéarité dans la séparation

entre les demi-espaces.

A partir de la représentation d'un neurone formel, il est possible de construire des réseaux qui combinent plusieurs neurones afin de modéliser des liens entre les entrées et la ou les sorties. Il existe beaucoup d'architectures de réseaux de neurones plus ou moins complexes. Nous décidons dans cette section de se limiter à la présentation de structures simples mais qui permettent de comprendre les concepts fondamentaux de ces structures.

4.2.5.2 Perceptron simple

Introduit en 1958 par Rosenblatt, le modèle du **perceptron** est une des architectures les plus simples de réseau de neurones [Rosenblatt1958]. Plusieurs structures dérivent de ce concept mais nous commençons par présenter la plus élémentaire : le perceptron simple à une couche cachée.

Structure du perceptron

La structure du **perceptron** à une couche cachée est décrite en Figure 4.14. Chacun des neurones Z présents dans la couche cachée sont composés d'une fonction de combinaison et d'une fonction d'activation, comme indiqué précédemment.

Dans le cas d'un perceptron simple à seuil, la sortie o_j du neurone j est donnée par le résultat de la fonction de Heaviside, de paramètre seuil θ_j , appliquée à la somme des produits termes à termes des composantes du vecteur d'entrée $\vec{x} = (x_1, x_2, \dots, x_N)$ et du vecteur de poids du neurone $\vec{w}_j = (w_{j1}, w_{j2}, \dots, w_{jN})$:

$$o_j = \begin{cases} 1 & \text{si } \sum_{i=1}^N w_{ji}x_i > \theta_j, \\ 0 & \text{sinon} \end{cases} \quad (4.22)$$

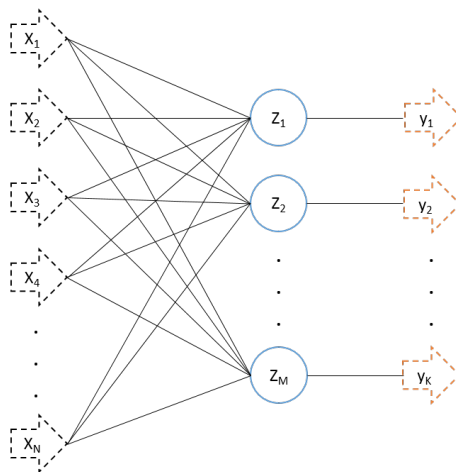


Figure 4.14 – Structure d'un perceptron simple, avec N entrées x , une couche cachée de M neurones Z et K les sorties y

L'entraînement de ce type de réseaux se fait le plus communément en faisant circuler les données d'entraînement, une par une ou par paquets (ou *batch*), de manière répétitive à travers le

réseau, jusqu'à obtenir un modèle convergent. Les poids des neurones sont modifiés à chaque itération pour améliorer la performance du réseau, c'est à dire diminuer l'erreur de prédiction. Dans le cas du perceptron, l'apprentissage se fait généralement par *propagation avant*, c'est à dire que l'information ne se déplace que des entrées vers les sorties. D'autres méthodes introduisent une récurrence, c'est à dire que les sorties sont utilisées pour alimenter les entrées, mais elle sont utilisées pour des modèles plus complexes de réseaux.

Plusieurs algorithmes d'apprentissage classiques pour un perceptron simple sont présentées par la suite.

Apprentissage par correction d'erreur

Dans le cas du perceptron simple à seuil, une des techniques les plus simples est l'apprentissage par *correction d'erreur*, dont l'algorithme est le suivant :

Apprentissage par correction d'erreur :

Entrées : un ensemble S de données d'apprentissages $(\vec{x}, y)$

- **Initialiser** les poids w des neurones aléatoirement.
- **Répéter** pour tous les éléments de l'ensemble d'apprentissage :
 - Prendre un exemple $(\vec{x}, y)$ des données d'apprentissage.
 - Calculer la sortie o pour l'entrée $\vec{x}$
 - **Pour** i de 0 à N :
 - $w_i \leftarrow w_i + (y - o)x_i$
 - **finPour**
- **finRépéter**

Sorties : Vecteur $\vec{w}_j = (w_{j1}, w_{j2}, \dots, w_{jN})$

On remarque que les valeurs des poids sont modifiées uniquement lorsque la sortie théorique y et la sortie calculée o sont différentes, d'où le terme de correction d'erreur. Si $o = 0$ et $y = 1$, l'algorithme augmente la force des poids, c'est le *renforcement* synaptique. Inversement, si $o = 1$ et $y = 0$, l'algorithme diminue la force des poids, c'est l'*inhibition* synaptique.

Les données d'apprentissage sont présentées de manière complète plusieurs fois jusqu'à ce qu'une présentation complète n'entraîne plus aucune modification dans les valeurs des poids. Il s'agit du critère de convergence de l'algorithme.

Si elle est simple à mettre en œuvre, cette technique d'apprentissage présente quelques inconvénients majeurs.

Premièrement, elle atteint une convergence seulement si l'échantillon d'apprentissage est linéairement séparable. Dans le cas où celui-ci ne l'est pas, il est théoriquement possible d'imaginer de stopper l'apprentissage quand l'algorithme *tourne en rond*, mais cela est compliqué en pratique de part la durée et la taille du calcul nécessaire. Cet algorithme est donc très sensible à des données aberrantes en apprentissage.

Deuxièmement, même si l'échantillon est linéairement séparable et que l'apprentissage converge, la solution retenue n'est pas optimisée pour être la plus *robuste*, c'est-à-dire la plus à même de généraliser sur des éléments nouveaux.

Apprentissage par descente de gradient

Il existe d'autres algorithmes d'apprentissage permettant de rechercher des modèles plus robustes et en meilleur capacité de gérer des ensembles de données non séparables linéairement. C'est le cas de l'apprentissage par descente de gradient, introduit en 1985 par Rumelhart [Rumelhart1985], et très utilisé dans l'apprentissage automatique. Le principe de cet algorithme est de chercher à minimiser une fonction d'erreur au lieu de chercher à classifier correctement toutes les données d'apprentissage.

La fonction d'erreur la plus communément utilisée pour cette application est l'erreur quadratique, définie pour perceptron par l'équation 4.23. Celle-ci est calculée directement sur la valeur du potentiel post-synaptique et non plus sur une sortie seuillée comme pour l'algorithme précédent. La valeur de sortie est donc réelle et non plus entière binaire.

$$E(\vec{w}) = \frac{1}{2} \sum_{i=1}^N (y_i - o_i)^2 \quad (4.23)$$

avec y_k la sortie attendue pour l'entrée $\vec{x}_k$ et o_k la sortie obtenue avec le vecteur $\vec{w}$ de poids testé.

Le but de la descente de gradient est de trouver un vecteur $\vec{w}$ tel que E est minimale, c'est à dire que le nombre d'erreur de prédiction sur l'apprentissage est minimal. La valeur particulière $E(\vec{w}) = 0$ caractérise le cas où le réseau ne fait aucune erreur sur la prédiction de l'ensemble S d'apprentissage.

La *descente* de gradient consiste alors à faire varier la valeur de chaque poids w_i d'une quantité Δw_i équivalente à la dérivée de $E(\vec{w})$ par rapport à w_i , c'est-à-dire le gradient de $E(\vec{w})$. Dans le cas du perceptron, il est possible de montrer que ce gradient vaut :

$$\frac{\partial E(\vec{w})}{\partial w_i} = - \sum_{i=1}^N (y_i - o_i) x_i \quad (4.24)$$

Les poids w_i sont modifiés d'une quantité Δw_i après une présentation complète de l'ensemble S des éléments d'apprentissage. Cette quantité est définie par l'équation 4.25.

$$\Delta w_i = -\eta \times \frac{\partial E(\vec{w})}{\partial w_i} = \eta \sum_{i=1}^N (y_i - o_i) x_i \quad (4.25)$$

où le coefficient η est le *taux d'apprentissage*, qui permet de régler le pas de la descente.

L'algorithme complet de la descente de gradient est alors :

Apprentissage par descente de gradient :

Entrées : un ensemble S de données d'apprentissages $(\vec{x}, y)$, η

- **Initialiser** les poids w des neurones aléatoirement.
- **Répéter**
 - **Pour** i de 0 à N : $\Delta w_i \leftarrow 0$ **finPour**
 - **Pour** tout $(\vec{x}, y)$ de S :
 - Calculer la sortie o pour l'entrée $\vec{x}$
 - **Pour** i de 0 à N :
 - $\Delta w_i \leftarrow \Delta w_i + \eta(y - o)x_i$
 - **finPour**
 - **finPour**
 - **Pour** i de 0 à N :
 - $w_i \leftarrow w_i + \Delta w_i$
 - **finPour**
- **finRépéter**

Sorties : Vecteur $\vec{w}_j = (w_{j1}, w_{j2}, \dots, w_{jN})$

Le taux d'apprentissage η est généralement choisi de manière empirique et suffisamment petit pour permettre à l'algorithme de converger vers le minimum de la fonction d'erreur. Néanmoins, si il est trop petit, le nombre d'itération sera très grand. Si il est trop fort, l'algorithme risque d'osciller autour du minimum sans jamais l'atteindre. Il est alors usuel de faire varier ce coefficient au cours de l'apprentissage, pour qu'il diminue graduellement à mesure que le minimum est approché.

Si cet algorithme présente une plus grande robustesse et une meilleure adaptation à des problèmes non linéairement séparables que le précédent, il est cependant lent à s'adapter aux données d'apprentissages car la modification des poids se fait par rapport à la globalité des éléments.

Pour terminer ce tour des méthodes d'apprentissages classiques pour un perceptron simple, l'algorithme de Widrow-Hoff, qui permet en général de converger plus rapidement vers une solution que la descente de gradient, est présenté par la suite.

Apprentissage de Widrow-Hoff

L'**apprentissage de Widrow-Hoff** est issu de la *règle de Widrow-Hoff* ou *règle delta*, introduite en 1960 par Widrow et Hoff [Widrow1960]. Cet apprentissage peut être considéré comme une variante de la descente de gradient et elle s'en différencie par sa règle de modification des poids. En effet, au lieu de modifier les poids par rapport à la globalité des éléments, ces derniers évoluent après la présentation de chaque élément. La règle de modification 4.25 est remplacée par la règle 4.26.

$$\Delta w_i = \eta(y_i - o_i)x_i \quad (4.26)$$

L'algorithme d'apprentissage est alors le même que pour la correction d'erreur, sauf pour la mise à jour des poids qui est réalisée via l'équation 4.26. De plus, étant donné que la sortie du perceptron est réelle pour l'apprentissage de Widrow-Hoff, les poids sont modifiés quasiment à chaque fois, ce qui n'est pas le cas avec la méthode par correction d'erreur.

L'avantage majeur de cet algorithme par rapport à celui de correction d'erreur est qu'il converge vers une solution même si l'ensemble des données d'apprentissage n'est pas linéairement séparable. Il est donc moins sensible à des données aberrantes.

4.2.5.3 Perceptron multi-couche

Une des faiblesses majeure des perceptrons simples est qu'ils ont des difficultés à traiter des problèmes non linéairement séparables. L'idée de mettre bout à bout plusieurs structures de perceptrons élémentaires a permis de venir à bout de ce problème. Les réseaux multi-couches se sont ainsi développés, notamment suite aux avancées de Le Cun et ensuite Rumelhart à propos de l'apprentissage par rétropropagation, en 1986 [Le Cun1986] [Rumelhart1986].

Structure d'un perceptron multi-couches

La structure d'un perceptron multi-couches est illustrée en Figure 4.15. Il est composé d'une couche de variables d'entrée, puis d'un certain nombre de couches de neurones cachés, et enfin d'une couche de neurones de sortie. Les entrées des neurones de chacune des couches cachées plus *profondes* que la première ainsi que celles de la couche de sortie sont reliées aux sorties des neurones de la couche précédente.

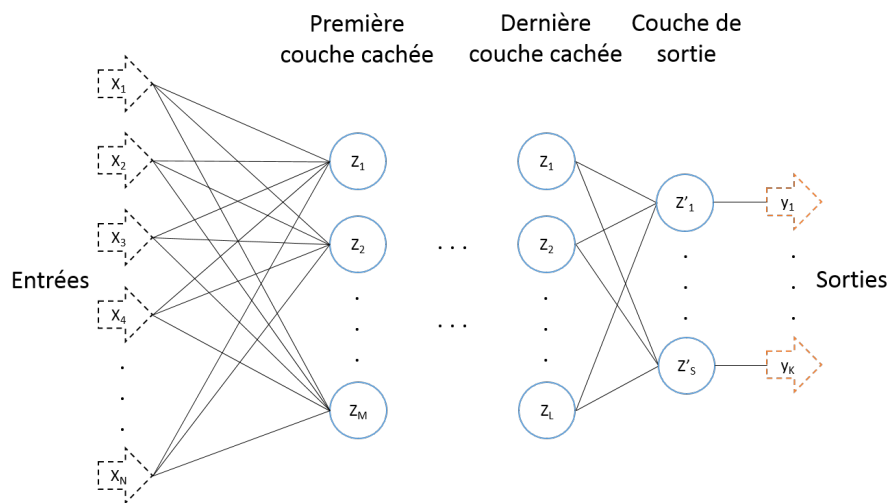


Figure 4.15 – Structure d'un perceptron multi couches

Une des difficultés principales avec ce type de réseau apparaît dans le choix du nombre de couches cachées ou encore du nombre de neurones par couche cachée pour construire un modèle adapté à un problème posé. Si certains algorithmes dits de *co-construction* existent pour tenter de faire ces choix de manière automatisée, ceux-ci sont généralement réalisés de manière empirique dans la pratique.

Afin d'entraîner ce type de système, une nouvelle méthode d'apprentissage a également été mise au point : la rétropropagation de gradient. Ses principes sont décrits dans la section suivante.

Apprentissage par rétropropagation de gradient

Le principe de cet algorithme est, comme celui de descente de gradient, de minimiser une fonction d'erreur caractérisant la capacité du modèle à bien prédire les éléments d'apprentissage. La difficulté supplémentaire dans le cas des réseaux multi-couches est de prendre en compte l'influence d'un changement de valeur d'un poids sur les neurones des couches suivantes.

Comme pour l'apprentissage de Widrow-Hoff, les poids des neurones sont modifiés à chaque passage d'un exemple d'apprentissage, d'une quantité Δw_{ij} .

Soit w_{ij} le poids associé au lien entre un neurone j vers un neurone i de deux couches successives, alors il est possible de définir $net_i = \sum_{Tot} w_{ij}x_{ij}$ comme le potentiel synaptique total du neurone i . On introduit alors le gradient d'erreur $\frac{\partial E(\vec{w})}{\partial net_i}$ par rapport à ce potentiel. On peut alors réécrire le gradient par rapport au poids w_{ij} comme :

$$\frac{\partial E(\vec{w})}{\partial w_{ij}} = \frac{\partial E(\vec{w})}{\partial net_i} \times \frac{\partial net_i}{\partial w_{ij}} = \frac{\partial E(\vec{w})}{\partial net_i} \times x_{ij} \quad (4.27)$$

En notant δ_i la quantité $-\frac{\partial E(\vec{w})}{\partial net_i}$, la variation de poids à appliquer, définie en équation 4.25, devient :

$$\Delta w_{ij} = \eta \times x_{ij} \times \delta_i \quad (4.28)$$

La quantité δ_i est définie différemment en fonction de si i est un neurone d'une couche cachée, ou de la couche de sortie. Il est possible de montrer que δ_i est définie respectivement dans chaque cas par les équations 4.29 et 4.30 .

Neurone caché :

$$\delta_i = o_i \times (1 - o_i) \times \sum_{l \text{ in } L} \delta_l w_{li} \quad (4.29)$$

avec o_i la sortie calculée du neurone i et L le nombre de neurones de la couche suivante.

Neurone de sortie :

$$\delta_i = o_i \times (1 - o_i) \times (y_i - o_i) \quad (4.30)$$

avec y_i la sortie théorique.

Les premiers calculs exécutés portent donc sur les poids de la couche de fin, et remontent ensuite à travers les couches cachées jusqu'à la couche d'entrée d'où le nom d'apprentissage par *rétropropagation*. En effet, les calculs des variations de poids pour les couches cachées prennent en compte les poids des couches suivantes dans l'ordre du réseau. Seule la couche de sortie peut être calculée directement avec les valeurs de sortie théorique d'apprentissage.

Apprentissage par rétropropagation de gradient :

Entrées : un ensemble S de données d'apprentissages $(\vec{x}, y)$, η

- **Initialiser** les poids w des neurones aléatoirement.
- **Répéter**
 - Calculer la sortie o pour une entrée $\vec{x}$ de S :
 - **Pour** tout neurone i de la couche de sortie :
 - $\delta_i = o_i \times (1 - o_i) \times (y_i - o_i)$
 - **finPour**
 - **Pour** toute couche cachée en remontant vers les entrées :
 - **Pour** tout neurone i de la couche courante et neurone l de la couche suivante :
 - $\delta_i = o_i \times (1 - o_i) \times \sum_{l \text{ in } L} \delta_l w_{li}$
 - **finPour**
 - **finPour**
 - **Pour** tout poids w_{ij} : $w_{ij} \leftarrow w_{ij} + \eta \times x_{ij} \times \delta_i$ **finPour**
- **finRépéter**

Sorties : Vecteurs $\vec{w}_{ij}$ des poids des différentes couches du réseau

Dans la pratique, le critère d'arrêt est souvent fixé comme un nombre d'itérations maximum du **Répéter** ou une valeur seuil basse à dépasser pour l'erreur. Le problème de sur-apprentissage se pose toujours et il est là aussi commun d'utiliser un lot de données de contrôle pour tester l'erreur de classification sur des données hors entraînement afin de valider le modèle.

Néanmoins, les réseaux multi-couches restent des modèles complexes à optimiser et il n'existe pas vraiment de méthode pour les faire converger à coup sûr.

Astuces pour améliorer la convergence

Si il n'existe pas encore de méthodes génériques pour assurer la convergence de l'algorithme, quelques conseils pratiques ont été donnés par LeCun dans un article très pédagogique [LeCun1998]. Les principales **astuces** pour "*améliorer grandement les chances de trouver une bonne solution tout en diminuant la vitesse de convergence*" portent sur :

- ⇒ Le choix entre l'*apprentissage par paquets* ou *batch learning* et l'*apprentissage stochastique* : le premier réfère à l'utilisation d'un lot d'éléments d'apprentissage à chaque itération et donc d'un gradient moyenné comme dans la descente de gradient classique, alors que le second n'utilise qu'un exemple par itération comme dans l'algorithme de Widrow-Hoff. Si l'apprentissage stochastique converge plus rapidement que l'apprentissage par paquets, ce dernier, notamment avec une taille de paquets adaptative, présente l'avantage d'être moins sensible aux minimas locaux.
- ⇒ Le *mélange des éléments de l'ensemble d'apprentissage*, ou leur présentation de manière aléatoire lors de l'apprentissage : les réseaux apprennent plus rapidement à partir d'exemple inattendus. Présenter des éléments successifs de classes différentes lors de l'apprentissage tend donc à améliorer la vitesse de convergence car ils contiennent une information non redondante.

- ⇒ La *normalisation des entrées* : la convergence est généralement plus rapide si la moyenne de chaque donnée d'entrée est proche de zéro et si leur covariance est à peu près la même. De plus, des données d'entrée décorréliées entre elles mènent également à une meilleur convergence.
- ⇒ L'*initialisation des poids* : dans le cas d'une fonction d'activation sigmoïde, les poids devraient être initialisés dans la région linéaire de cette fonction, c'est à dire dans un intervalle de valeurs autour de l'origine. Néanmoins, les poids ne doivent pas être trop faible au risque d'obtenir une convergence lente. Pour un ensemble de données normalisées et une fonction d'activation sigmoïde approximée et de tangentes $[-1, 5; 1, 5]$, LeCun conseille d'initialiser les poids d'un neurone aléatoirement suivant une distribution uniforme de moyenne 0 et d'écart type $m^{-1/2}$, avec m le nombre d'entrée du neurone.
- ⇒ La valeur du *taux d'apprentissage* : LeCun recommande de donner des taux d'apprentissage différents à chaque poids pour améliorer la convergence. Le but étant de faire converger les poids du réseau complet à la même vitesse. De plus, comme déjà abordé précédemment, le fait de choisir des taux adaptatifs, c'est-à-dire qui évolue au cours de l'apprentissage aurait également un effet positif sur la convergence. Il est usuel de faire décroître ce taux à mesure que la fonction d'erreur approche d'un minimum.
- ⇒ L'introduction d'un nouveau terme appelé *moment* ou *momentum* : le moment intervient dans la règle de modification des poids en prenant en compte les modifications antérieures pour éviter de rester dans un minimum local. La nouvelle règle de modification des poids est définie en équation 4.31.

$$\Delta w_i(t) = \eta \times x_{ij} \times \delta_i + \mu \times \Delta w_{ij}(t-1) \quad (4.31)$$

avec t le numéro de l'itération actuelle et μ le coefficient attaché à la force du moment.

4.2.5.4 Forces et limites

Quelques forces et limites des réseaux de neurones sont présentées, sur la base des concepts exposés précédemment.

Parmi les atouts majeurs de ces techniques, on retrouve :

- ⇒ Leur capacité à modéliser des problèmes linéaires ou non-linéaires, autant simples que complexes, grâce aux réseaux multi-couches.
- ⇒ Leur capacité à traiter des problèmes multi-classes.
- ⇒ Leur résistance aux exemples d'apprentissage aberrants grâce à des techniques d'apprentissage itératives.
- ⇒ L'inituitivité de leur entraînement, qui requiert moins de compréhension statistiques que les autres méthodes d'apprentissage automatique. Un réseau apprend de lui-même en ajustant ses poids à force de passer en revue les données d'apprentissage.

Parmi les difficultés majeures de ces techniques, on retrouve :

- ⇒ La difficulté à trouver une architecture optimale en terme de nombre de couches et de neurones. Il n'existe pas de méthode systématique et le choix est souvent basé sur des tests empiriques.
- ⇒ La difficulté à optimiser le réseau, une fois l'architecture choisie, pour garantir une convergence. Même si des astuces existent (cf. Section précédente), la sélection automatique de paramètres optimaux reste encore aujourd'hui un enjeu majeur de la recherche sur ces algorithmes.

⇒ La difficulté d'interprétation de la décision à travers le modèle final. Comme les SVM, il s'agit d'une *boîte noire* dont le parcours pour atteindre un résultat donné est difficile à analyser. L'interprétabilité des résultats de tels algorithmes est également un enjeu de recherche actuel.

4.2.6 Algorithme des k-moyennes

L'algorithme des *k-moyennes*, ou *k-means* en anglais fait parti de la catégorie des méthodes de classification non-supervisées. L'algorithme est également parfois désigné par la "méthode Llyod-Forgy", du nom de ses premiers développeurs [Forgy1965]. Cette techniques a ensuite été popularisée pour des applications de classification par MacQueen en 1967 [MacQueen1967]. Nous proposons de présenter brièvement le principe de fonctionnement de cette technique dans cette section.

4.2.6.1 Principe de fonctionnement

La méthode des k-moyennes est en réalité une méthode de *clustering*, c'est-à-dire de rassemblement de données par groupes en cherchant à minimiser une certaine fonction. On parle aussi de *partitionnement* de données.

Dans le cadre classique, la méthode fixe le centre d'un groupe comme le barycentre des points de ce groupe, c'est-à-dire la moyenne des positions de chaque point. La fonction à minimiser pour optimiser les groupements est alors la somme des carrés des distances des points aux centres de leurs groupes respectifs. La distance choisie est classiquement la distance Euclidienne, mais d'autres distances peuvent être utilisées, comme celles présentées dans l'algorithme des k-NN (Tableau 3.2).

Le fonctionnement de l'algorithme, illustré en Figure 4.16, est le suivant :

Algorithme des K-moyennes :

Entrées : un ensemble S de données, k

- **Initialiser** les positions des k centres
- **Répéter** jusqu'à convergence
 - **Assigner** les points de l'ensemble S aux centres les plus proches
 - **Calcul les barycentres** des groupements
- **finRépéter**

Sorties : Groupements des données selon k classes

En ce qui concerne le *critère d'arrêt*, il peut être défini de différentes manières. L'algorithme peut s'arrêter si plus aucun changement n'est observé dans l'assignation des données aux k groupes entre deux itérations, il y alors convergence. Dans le cas où le temps de calcul est trop long, l'algorithme peut être stoppé après un certain nombre d'itérations, fixé en avance.

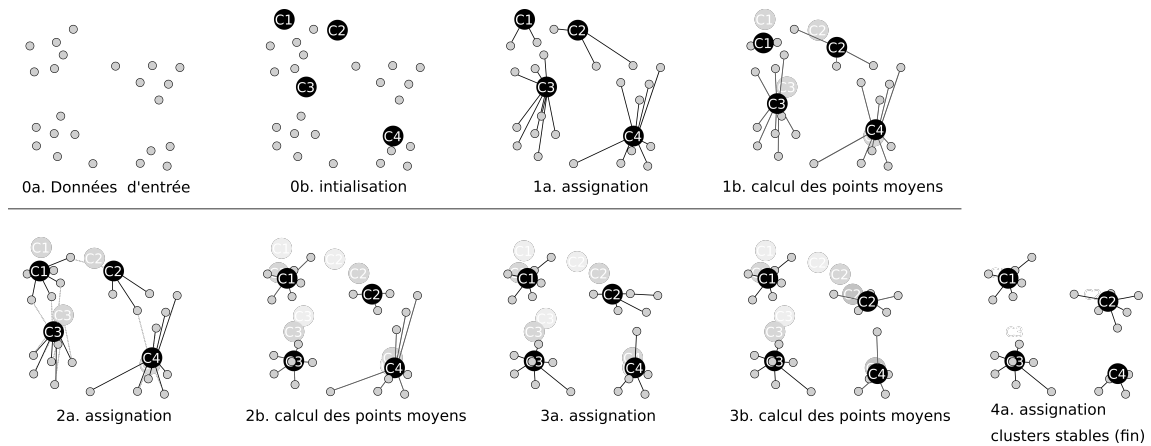


Figure 4.16 – Principe et étapes de fonctionnement de l'algorithme des K-moyennes- *By Mquantin - Own work, CC BY-SA 4.0*

L'initialisation des centres au début de l'algorithme peut également être réalisée de différentes manières. Par exemple, dans la méthode initiale, Forgy décide d'assigner les k centres à k données d'entrée aléatoirement choisies. Une autre méthode propose quant à elle de commencer par assigner aléatoirement un groupe à chaque donnée d'entrée, et de commencer directement par le calcul des barycentres des points de chaque groupe (deuxième étape de la boucle itérative de l'algorithme). Le résultat donné par la suite de l'algorithme est très sensible à l'initialisation des centres. En effet certaines configurations initiales qui peuvent mener à un minimum d'optimisation local, comme illustré en Figure 4.17. Il est donc usuel de tester plusieurs configurations initiales et de choisir ensuite le meilleur résultat.

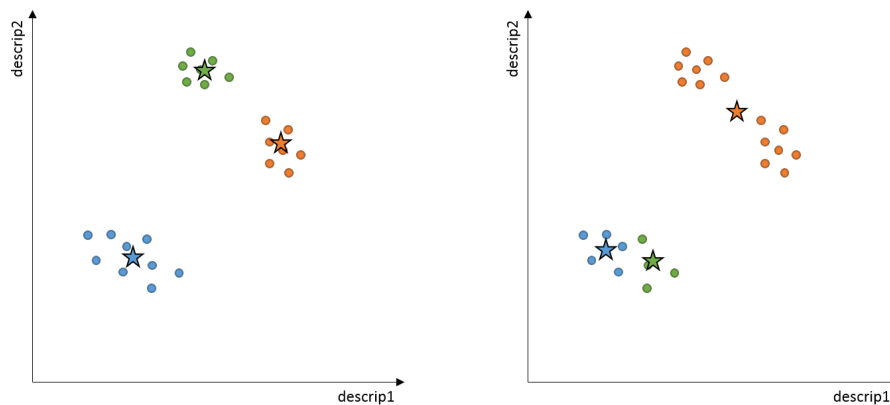


Figure 4.17 – Exemple de deux exécutions sur un même ensemble de données, menant à deux partitionnements différents

La méthode des k-means++, introduite en 2007 par Arthur et Vassilvitskii [Arthur2007], propose une autre méthodologie pour initialiser les centres, permettant à la fois d'améliorer grandement l'erreur de classification finale, c'est-à-dire la convergence vers de meilleures solutions, mais également la vitesse de traitement. Cette méthode part du principe que l'algorithme des k-moyennes fonctionne mieux si les centres sont répartis de manière plus écartée dans l'espace des données au départ.

Cet algorithme consiste en la décomposition de l'initialisation des centres en 3 étapes :

1. Choisir un centre c_1 de manière aléatoire et uniforme parmi les données x de l'ensemble S

2. Prendre un nouveau centre c_i , de manière aléatoire parmi les données x de l'ensemble S , mais avec une densité de probabilité $\frac{D(x)^2}{\sum_{x \in X} D(x)^2} \cdot D(x)$ correspondant à la plus petite distance entre x et le centre le plus proche ayant déjà été choisi précédemment.
3. Répéter l'étape 2 jusqu'à ce que k centres aient été choisis.

La suite de l'algorithme suit la méthode des k-moyennes classiques.

4.2.6.2 Forces et limites

Quelques forces et limites de l'algorithme des k-moyennes sont présentées, sur la base des concepts exposés précédemment.

Parmi les atouts majeurs de cette technique, on retrouve :

- ⇒ Son fonctionnement intuitif et sa simplicité de mise en œuvre.
- ⇒ Sa vitesse de traitement, par rapport aux autres algorithmes de partitionnement, en gardant k faible.

Parmi les difficultés majeures de cette technique, on retrouve :

- ⇒ La difficulté du choix de k sans connaissance *a priori*. Les résultats donnés par l'algorithme peuvent être très différents d'un choix de k à un autre.
Une solution peut alors être de tester plusieurs valeurs de k ou de visualiser les données au préalable, via une ACP par exemple, pour obtenir une information *a priori* guidant le choix. Néanmoins, le choix de k peut également être un avantage dans certaines applications pour lesquelles le nombre de classes est fixé par avance. C'est le cas de la représentation de textes ou d'images par sacs de mots, dans laquelle un dictionnaire de taille k correspondant à k mots d'intérêts est fixé par avance.
- ⇒ La dépendance de la convergence par rapport à l'initialisation des centres, comme illustrée en Figure 4.17, qui induit un risque d'obtenir une solution sous-optimale. Certaines techniques pour améliorer ce point ont été décrites précédemment.
- ⇒ Si l'algorithme est avantageux en temps de calcul pour un petit nombre de classes k , le temps de traitement augmente tout de même rapidement avec le nombre de données dans S , le nombre de dimensions des données et le nombre de classes k .

4.2.7 Réseaux de Deep Learning

Pour terminer ce tour d'horizon des principales techniques de classification, nous souhaitons introduire une branche des réseaux de neurones ayant pris de plus en plus d'ampleur lors de ces dernières années dans les domaines de l'apprentissage automatisé et de l'intelligence artificielle. Il s'agit des techniques d'*Apprentissage Profond*, mieux connues sous le terme anglophone de *Deep Learning*. La logique de fonctionnement de ces algorithmes diffère un peu de celle des méthodes présentées jusqu'ici, c'est pour cela que nous la présentons en dernier.

4.2.7.1 Fondements historiques

Bien que les concepts fondant le Deep Learning datent de la même époque que le développement des premiers réseaux de neurones classiques (années 1960), le véritable essor de ces techniques a eu lieu bien plus récemment. En effet, le développement de nouvelles architectures conjugué aux progrès grandissant des performances des machines de calcul ont en grande partie permis de rendre ces algorithmes fonctionnels. Les travaux de Y.LeCun à la fin des années 1980/début des années 1990 à propos de la reconnaissance de chiffres manuscrits ont notamment permis de

populariser cette approche [LeCun1989] [LeCun1990].

Les architectures de Deep Learning proposées depuis et leurs applications sont très nombreuses tant le domaine est en pleine croissance. Ne souhaitant aborder ici seulement les principes fondamentaux de ces réseaux, nous proposons au lecteur de se référer aux récents et très complets articles de LeCun et Schmidhuber pour des exposés plus complets sur le fonctionnement et l'histoire de ces techniques [LeCun2015] [Schmidhuber2015].

Principes de fonctionnement

Les techniques d'apprentissage automatiques classiques telles que celles présentées jusqu'à maintenant se fondent sur l'exploitation d'une information simplifiée. En effet, les données brutes, telles que des images par exemple, ne sont pas directement utilisées pour l'entraînement des algorithmes. Elles sont tout d'abord simplifiées sous forme de descripteurs caractéristiques (cf. Chapitre 2). Il est donc nécessaire de bien concevoir la description des images afin de fournir une information pertinente à l'algorithme d'apprentissage.

Les réseaux de Deep Learning ont eux pour vocation de s'entraîner directement sur les données brutes, en décrivant par eux même les données au cours du processus d'apprentissage. Ils sont en effet capables, dans une certaine mesure, de trouver par eux-même les représentations pertinentes pour effectuer des tâches de détection ou de classification. Les données initiales sont transformées en des représentations de plus hauts niveaux par un enchaînement de modules de transformations non-linéaires simples. Ces différentes représentations sont de plus en plus abstraites au fur et à mesure des traitements, mais elles concentrent l'information pertinente pour effectuer une tâche donnée. Ainsi, en combinant assez de transformations, ces modèles sont capables d'apprendre des fonctions très complexes.

De ce fait, les réseaux de neurones en apprentissage profond sont composés de beaucoup plus de couches et de neurones que les réseaux classiques.

Différentes architectures de réseaux existent, mais nous choisissons de ne présenter ici que les réseaux de convolutions (*CNN* pour Convolutional Neural Network) qui sont les plus communément utilisés. Un exemple d'architecture d'un tel réseau est exposé en Figure 4.18. Parmi les modules de transformations basiques le constituant, on retrouve :

- Les **couches de convolution** :

La convolution consiste à appliquer des filtres aux données d'entrée afin d'en retenir les caractéristiques pertinentes. Contrairement à l'apprentissage automatique classique dans lequel ces filtres sont définis par avance lors de la conception du système, les couches de convolution en Deep Learning apprennent automatiquement ces filtres au fur et à mesure de l'entraînement. Chaque couche contient un certain nombre de filtres, alors nommés *noyaux de convolution*, dont les sorties sont appelées *cartes de caractéristiques* ou *feature maps*.

La sortie d'une couche de convolution comportera donc un plus grand nombre de cartes qu'en entrée, mais de taille plus faible.

- Les **couches de pooling/subsampling** :

Les modules de pooling permettent de réduire la taille des cartes de caractéristiques tout en gardant l'information importante contenue dans celles-ci. Le processus consiste simplement à faire glisser une fenêtre dans la carte de départ et de récupérer une valeur caractéristique à chaque pas. La fenêtre de valeurs dans la carte de départ sera donc représentée par une seule valeur dans la carte d'arrivée. Cette valeur peut être par exemple la moyenne, on parle alors de *mean pooling*, ou encore le maximum, on parle alors de *max pooling*. En pratique, c'est le maximum qui est généralement utilisée car il fonctionne bien empiriquement.

La sortie d'une couche de pooling consiste donc en un nombre identique de cartes qu'en entrée, mais de tailles plus petites. Cela permet de faciliter les opérations qui suivent.

— Les **couches de rectification** :

Même si elle n'apparaissent pas sur l'exemple proposé, ces couches sont communément utilisées dans les CNN. Elles consistent en l'application de fonctions d'activations sur des cartes de caractéristiques. La plus utilisée en Deep Learning est la fonction ReLU (Unité Linéaire Rectifiée) qui a pour effet de ramener à 0 toutes les valeurs négatives d'une carte. Elle a pour effet d'améliorer la capacité du modèle à prendre en compte les effets d'interactions entre variables mais aussi les effets non-linéaires.

La sortie d'une couche de rectification comporte donc le même nombre de cartes qu'en entrée, de même tailles, mais avec ses valeurs rectifiées.

— Les **couches entièrement connectées (fully connected)** :

Les couches entièrement connectées fonctionnent comme les couches de neurones des réseaux de neurones classiques présentés précédemment. Chaque neurone de cette couche est lié à toutes les sorties des neurones de la couche précédente et transforme les cartes en une liste de votes qui peuvent par exemple être utilisés pour classifier les données d'entrée du réseau.

D'autres types de couches existent et entrent en jeu dans des architectures de réseaux plus compliquées, par exemple les couches de perte ou *dropout* qui permettent de rendre l'entraînement du réseau plus robuste. Dans le cadre de cette section, nous choisissons de nous limiter aux éléments de bases décrits précédemment qui permettent de comprendre globalement les principes de l'apprentissage profond.

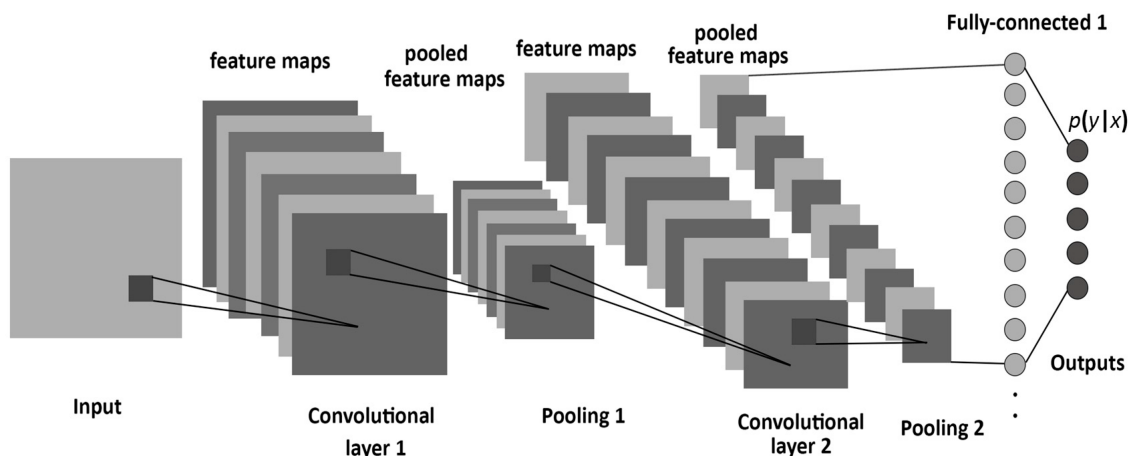


Figure 4.18 – Architecture typique d'un réseau de convolution [Albelwi2017]

4.2.7.2 Performances d'un réseau de Deep Learning

Les performances d'un réseau de Deep Learning peuvent être influencées par de nombreux choix dans la construction et l'entraînement de celui-ci.

Certaines problématiques qui se posaient pour les réseaux perceptron multi-couches classiques se posent également en apprentissage profond. C'est par exemple le cas du choix du taux d'apprentissage, pour lequel les astuces valables dans le cas des réseaux standards prévalent également pour le Deep Learning.

Les choix du nombre, de la forme et de la taille des filtres utilisés dans les couches de convolution mais également de la taille des fenêtres de pooling jouent également un rôle important dans les performances de ce type de réseaux. Ces choix dépendent notamment du type et de la taille des données en entrée, de la quantité de données disponibles ou encore du type et de la complexité du

traitement réalisé.

De plus, le risque de *surapprentissage* est très fort pour ce type d'architecture de part notamment leur taille et le nombre de traitements réalisés.

La première implication de cette observation est que les réseaux de Deep Learning requièrent en général un très grand nombre de données d'apprentissage pour être entraînés efficacement et éviter le surapprentissage. Néanmoins, il existe des méthodes pour limiter ce phénomène dont certaines sont décrites ci-après.

Tout d'abord, du point de vue de l'architecture du réseau, la limitation de la *taille du réseau* est la technique la plus simple à mettre en œuvre pour réduire le risque de surapprentissage. Néanmoins, plus le réseau est petit, plus il perd de sa puissance et de son potentiel de prédiction.

Toujours par rapport à l'architecture du réseau, il est possible d'utiliser des *couches de perte* ou *Dropout* et de *DropConnect* dont le rôle est de désactiver des neurones ou des connexions entre neurones aléatoirement à chaque itération d'apprentissage. Ceci a pour conséquence de permettre aux neurones de s'entraîner plus indépendamment les uns des autres, et ainsi de donner au réseau une meilleure capacité de généralisation. Au delà de l'architecture du réseau, il est également possible de jouer sur les données d'apprentissage. En effet, nous avons introduit précédemment le fait qu'un volume de données d'apprentissage trop faible peut également mener à du surapprentissage. Un tel manque peut être en parti pallié avec certaines techniques comme celle de l'*augmentation de données*. Cette technique, surtout utilisée avec des images, consiste en l'ajout de données artificielles au volume de données d'apprentissage, dérivant souvent des données existantes. L'ajout de bruit, l'application d'opérations de translation et rotation ou encore la modification de l'histogramme des images sont couramment utilisés. Il faut par contre faire attention à ce que l'augmentation de données aille dans le "bon sens", c'est-à-dire que les données modifiées restent plausibles dans l'application visée. Il peut en effet être contre productif d'entraîner un modèle sur des données non pertinentes.

Si nous nous limitons à la présentation de ces quelques astuces d'optimisation, nous souhaitons faire remarquer que de nombreuses autres méthodes ont été développées pour améliorer les performances des réseaux d'apprentissage profond.

C'est notamment le cas de l'utilisation de réseaux pré-entraînés afin de pouvoir utiliser cette méthode sur des petits lots de données d'apprentissage spécifique au problème abordé. Nous proposons de présenter les principes de cette technique dans la partie suivante.

4.2.7.3 Utilisation de réseaux pré-entraînés

L'utilisation de réseaux pré-entraînés est également référencée sous le nom d'*apprentissage par transfert* ou *transfer learning*. Ces techniques permettent de faire du Deep Learning en diminuant drastiquement les temps d'entraînement et le nombre de données d'apprentissage nécessaires. Le principe réside dans l'utilisation de connaissances construites par un réseau de neurone sur un problème donné afin de résoudre un autre problème plus ou moins similaire. On parle aussi de *transfert de connaissances*.

Il existe différentes stratégies permettant d'exploiter un réseau pré-entraîné :

- **Le fine-tuning total** : cette stratégie consiste à remplacer la dernière couche *fully-connected* du réseau pré-entraîné par un classifieur classique adapté au nouveau problème, tels que ceux présentés précédemment. Toutes les couches du réseau sont ensuite entraînées sur les nouvelles données.

Cette technique permet de réduire le temps d'apprentissage car les coefficients des différentes couches sont initialement ceux du réseau pré-entraîné. Néanmoins, le nombre de données d'apprentissage nécessaire reste grand car toutes les couches sont entraînées.

- **L'extraction de paramètres** : cette stratégie consiste à utiliser le réseau pré-entraîné comme générateur de descripteurs des données du nouveau problème. Ce cas est particulier car le réseau n'est pas utilisé comme classifieur, mais comme descripteur, à l'image de ceux présentés dans le Chapitre 2.

La dernière couche *fully-connected* est retirée et le réseau alors tronqué sert à extraire des représentation des données, qui sont ensuite utilisées pour entraîner un classifieur classique. Les coefficients des couches du réseau sont dans ce cas fixes et résultent de son entraînement antérieur. En d'autres termes, les couches du réseau ne sont pas ré-entraînées.

Il est néanmoins généralement nécessaire que les nouvelles données décrites de cette manière soient proches des données d'entraînement de base du réseau tronqué, au risque de générer des représentations non-pertinentes.

- **Le fine-tuning partiel** : cette stratégie est à la jonction des deux stratégies précédentes. On fixe les coefficients de certaines couches du réseau pré-entraîné, et on remplace la couche *fully-connected* par un nouveau classifieur. Les nouvelles données servent alors à entraîner le nouveau classifieur et les coefficients des couches non fixées, qui correspondent généralement aux couches les plus profondes du réseau.

Cette stratégie a pour avantage de nécessiter moins de nouvelles données d'entraînement car toutes les couches ne sont pas ré-entraînées. De plus, cette stratégie peut fonctionner avec des nouvelles données d'entraînement très différentes de celles utilisées initialement pour le réseau pré-entraîné. Cette propriété semble venir du fait que les premières couches d'un réseau décrivent des représentations assez génériques, qui peuvent donc se retrouver dans des données différentes, alors que les dernières sont plus spécifiques au problème visé. Le fait de fixer les premières couches et de ré-entraîner les dernières permet donc de garder l'information extraite sur les données précédentes, tout en se spécialisant sur un nouveau problème.

Différents réseaux pré-entraînés sont disponibles en libre accès pour la communauté du Deep Learning. Cet ensemble de réseaux est communément appelé un *zoo de modèles* et regroupe des architectures entraînées pour de la reconnaissance d'images, reconnues dans la communauté des chercheurs et développeurs. Parmi les configurations les plus utilisées, il est possible de citer [Deeplearning4j2017] :

- | | |
|-------------|----------|
| • AlexNet | • LeNet |
| • DarkNet | • ResNet |
| • GoogLeNet | • YOLO |
| • Inception | • VGG |

Beaucoup de ces réseaux sont entraînés via la base d'images **ImageNet**. Elle contient à l'heure actuelle plus de 14 millions d'images annotées par un processus de production participative sur internet.

4.2.7.4 Forces et limites

Quelques forces et limites des réseaux de neurones sont présentées, sur la base des concepts exposés précédemment.

Parmi les atouts majeurs de ces techniques, on retrouve :

- ⇒ Leur capacité à construire eux-même les descripteurs pertinents pour le problème donné à résoudre au réseau. Cependant, l'optimisation des filtres de convolution utilisés pour construire ces descripteurs n'est pas évidente.

- ⇒ Leur hautes performances pour traiter des problèmes relevant de nombreuses applications dans des domaines très divers. Ainsi il est possible de retrouver des utilisations du Deep Learning allant de la description d'images ou de vidéos, à la reconnaissance de parole, en passant par la traduction linguistique de textes ou encore le pilotage de robots et voitures autonomes.

Les algorithmes de Deep Learning ont permis de faire des bons de performances dans certaines applications où les algorithmes classiques d'apprentissage automatique ne progressaient plus beaucoup. Ils jouent d'ailleurs un rôle fondateur dans les progrès importants que connaît l'intelligence artificielle à l'heure actuelle.

Parmi les difficultés majeures de ces techniques, on retrouve :

- ⇒ L'explosion des topologies de réseaux, en termes de nombre et de types de couches, qui rend très compliqué le choix d'une architecture. La mise au point de la topologie la plus pertinente pour une application donnée relève pour le moment d'un véritable problème de recherche.
- ⇒ Au delà de la topologie du réseau, l'optimisation des hyper-paramètres, plus nombreux que pour des réseaux perceptrons classiques, reste également compliquée. Ainsi, les choix de la forme et du nombre de filtres de convolution et de la taille des fenêtres de pooling sont encore généralement fondés sur des tests empiriques.
- ⇒ Le besoin de très grands lots de données d'apprentissage, qui est un des problèmes principaux pour le développement de ce type de réseaux sur certaines applications à l'heure actuelle, notamment en industrie.
- ⇒ Le temps d'entraînement de ce type de réseau, qui est souvent très long (plusieurs heures à plusieurs jours en fonction des topologies et données d'apprentissage). Il est néanmoins possible dans certains cas d'utiliser des réseaux pré-entraînés dont seulement les dernières couches sont ré-entraînés pour l'application visée, ce qui permet de réduire grandement le temps d'apprentissage.
- ⇒ Tout comme les réseaux de neurones classiques, ils sont considérés comme des boîtes noires car il est difficile d'interpréter le chemin suivi pour obtenir une certaine prédiction une fois le modèle entraîné. Néanmoins, l'ouverture de cette boîte noire pour mieux comprendre le fonctionnement profond de ces réseaux fait l'objet de plus en plus de travaux de recherches à l'heure actuelle.

4.3 Discussion quant aux caractéristiques des différentes méthodes et leur application industrielle

Après avoir présenté un aperçu du fonctionnement des principales familles d'algorithmes utilisées pour classer des données en apprentissage automatique, nous proposons de récapituler leurs caractéristiques respectives, ainsi que les avantages et inconvénients qui y sont associés. Nous exposons tout d'abord une comparaison des différents algorithmes par rapport à certains critères sélectionnés, puis nous présentons une mise en perspective de ces critères dans le contexte du contrôle d'aspect automatisé sur des lignes de production.

4.3.1 Comparaison des techniques

Pour effectuer cette comparaison, nous avons repris certaines conclusions exposées par Kotsiantis dans une comparaison entre techniques d'apprentissage non-supervisées [Kotsiantis2007], en les adaptant aux techniques que nous avons choisi de présenter, mais également en y ajoutant certains éléments de comparaison.

Nous avons par contre décidé d'inclure les réseaux de Deep Learning dans la comparaison, car même si ils présentent des différences notables avec les méthodes d'apprentissage automatique classiques, leur fonctionnement reste supervisé et les critères choisis peuvent donc s'appliquer à eux également. De plus, au vue du développement exponentiel récent de ces techniques dans le domaine de l'apprentissage automatisé et de l'intelligence artificielle, nous souhaitons les prendre en considération dans notre étude.

Le Tableau 4.1 résume le comparatif des algorithmes de classifications développé dans la suite de cette section. Nous avons choisi de comparer les méthodes présentées dans ce chapitre, à l'exception des k-moyennes. Faisant parti des techniques non-supervisées de rassemblement de données, il est en effet compliqué de la comparer aux autres suivants les mêmes critères.

Les critères d'intérêt que nous avons choisis sont les suivants :

- La **vitesse d'entraînement** : elle caractérise la vitesse de calcul de l'algorithme pour la phase d'apprentissage du modèle, indépendamment du nombre de données d'entraînement ou de la dimension de celles-ci ;
- La **vitesse de classification** : elle caractérise la vitesse de calcul de l'algorithme pour classer un nouvel élément, une fois le modèle entraîné ;
- L'**occupation de la mémoire après entraînement** : elle caractérise la place en mémoire occupé par le modèle entraîné. Si l'algorithme est performant sur ce critère, cela implique qu'il occupe une place faible ;
- La **robustesse aux données d'apprentissage aberrantes** : elle caractérise la tolérance de la méthode à la présence de données mal étiquetées dans les exemples d'apprentissage ;
- La **capacité d'entraînement avec peu d'exemples** : elle caractérise l'aptitude du modèle à entraîner un modèle performant sans un très grand nombre d'exemples ;
- La **robustesse au phénomène de surapprentissage** : elle caractérise le niveau de sensibilité de l'algorithme au risque de sur-apprentissage, c'est-à-dire de sur-spécialisation du modèle sur les données d'entraînement ;
- La **capacité à traiter des problèmes non-linéaires compliqués** : elle caractérise le potentiel de l'algorithme à modéliser des fonctions non-linéaires compliquées, c'est-à-dire sa capacité à garder une bonne performance pour des problèmes difficiles ;
- La **facilité d'optimisation** : elle caractérise la complexité d'optimisation des hyper-paramètres de l'algorithme pour obtenir un modèle fonctionnel et performant ;
- L'**interprétabilité de la classification** : elle caractérise le fait de pouvoir donner du sens ou non au chemin suivi par le modèle pour prendre une décision de classification. Si l'algorithme est performant pour ce critère, cela implique que les étapes ayant mené à un résultat sont facilement exploitables pour expliquer ce résultat ;
- L'**interprétabilité du fonctionnement** : elle caractérise le degré de difficulté à comprendre le principe de fonctionnement de l'algorithme, c'est à dire la manière de construire un modèle ;
- La **robustesse à la dimension des données** : elle caractérise la capacité du réseau à traiter des données de dimensions différentes, et notamment de grandes dimensions.

	k-NN	SVM	Arbres déci- sionnels	Forêts d'arbres	Réseaux baye- siens naifs	Réseaux de neu- rones	Deep Lear- ning
Vitesse d'entraînement	++++ (pas de phase expli- cite)	+	+++	++	++++	+	+
Vitesse de classification	+	++++	++++	++++	++++	++++	++++
Occupation mémoire après entraînement	+	++++	++++	++++	++++	++++	++++
Robustesse aux données aberrantes	++ (dépend de k)	+++ (marge souple)	+	+++	+++	+++	++++
Capacité d'entraînement avec peu d'exemples	+++	++	+++	+++	+++	++	+ (sauf réseaux pré- entraînés)
Facilité d'optimisation	+++	+	+++	+++	++++	+	+
Robustesse au surapprentissage	+++	++	++	++	+++	+	+
Capacité à traiter des problèmes compliqués	+	+++	+	+	+	++++	++++
Interprétabilité de la classification	++	+	++++ (sauf grands arbres)	++	++++	+	+
Interprétabilité du fonctionnement	++++	++	++++	++++	++++	++	+
Robustesse à la dimension des données	++	++++	++	++	+++	++++	++++

Tableau 4.1 – Comparaison des algorithmes de classification automatique (++++ très performant / + peu performant)

4.3. Discussion quant aux caractéristiques des différentes méthodes et leur application industrielle

Du point de vue de la vitesse d'entraînement, les algorithmes les plus simples tels que les arbres décisionnels ou les réseaux bayésiens naïfs, sont les plus performants, c'est-à-dire les plus rapides. Les forêts d'arbres sont quant à elles un peu plus lentes à entraîner car il s'agit d'apprendre plusieurs arbres et non un seul. Néanmoins, ces méthodes restent bien plus rapides que les algorithmes plus compliqués tels que les SVM et les réseaux de neurones. Les réseaux de neurones de Deep Learning peuvent prendre plusieurs heures voir plusieurs jours pour construire leur modèle car ils travaillent sur de très grands lots de données.

La méthode des k -NN ne possède pas de phase d'entraînement explicite car l'algorithme n'entraîne pas de modèle comme les autres techniques. Nous considérons donc qu'il est très performant par rapport aux autres sur ce critère de comparaison.

Néanmoins, pour ce qui est de la vitesse de classification et de l'occupation de la mémoire après entraînement, le statut de classifieur paresseux de l'algorithme des k -NN joue en sa défaveur. En effet, le fait de devoir garder en mémoire l'ensemble des données d'apprentissage et de répéter le processus complet de calcul de distances pour chaque classification rend cette étape très gourmande en temps de calcul et en place occupée.

Tous les autres algorithmes sont beaucoup plus performants sur ces deux points car ils ne stockent seulement qu'un modèle entraîné qui occupe bien moins d'espace que les données d'apprentissage. De plus, le temps de calcul pour réaliser une nouvelle classification une fois le modèle entraîné est souvent très faible.

De manière général, les algorithmes d'apprentissage itératifs comme ceux basés sur les réseaux de neurones sont robustes aux données aberrantes car l'effet de ces dernières est lissé au cours du processus.

Les réseaux bayésiens naïfs sont également assez résistants car leur principe consiste en l'apprentissage de lois de probabilités pour chaque descripteur, ce qui permet de limiter l'effet de données erronées.

La plupart des autres algorithmes sont sensibles à la présence d'exemples d'entraînement erronés de part leurs principes de fonctionnement. Néanmoins, des variantes ou méthodes d'optimisation existent pour chaque type de méthode afin de réduire cette sensibilité. Un choix de k grand pour les k -NN ou l'utilisation des variantes de la marge souple ou des forêts d'arbres aléatoires respectivement pour les SVM et les arbres décisionnel permettent de rendre ces algorithmes plus robuste sur ce critère.

Dans tous les cas il apparaît comme logique que le nombre de données aberrantes ne doit pas constituer une part trop importante des données d'apprentissage, au risque d'apprendre un modèle faussé.

La qualité des exemples d'entraînement guide la qualité de la classification quelque soit l'algorithme. Leur nombre est également un critère important pour permettre au modèle de bien généraliser sur des données de test ensuite. Les SVM et les réseaux de neurones ont généralement besoin de plus d'exemples que les autres algorithmes pour générer un modèle satisfaisant.

Ce problème s'applique particulièrement aux réseaux de neurones car plus ils sont grands et constitués de beaucoup de neurones, plus ils nécessitent de données d'entraînement pour éviter le surapprentissage. Dans ce contexte, les réseaux de Deep Learning requièrent en général des ensembles de données très volumineux (des milliers voir bien plus dans les réseaux les plus compliqués). Il existe néanmoins des techniques pour réduire le nombre de données d'apprentissage nécessaires comme l'utilisation de réseaux pré-entraînés.

Si le manque d'exemples peut mener au phénomène de surapprentissage, d'autres problèmes peuvent également y participer, comme la difficulté d'optimisation de certains hyper-paramètres. De manière général, plus il possède d'hyper-paramètres, plus un algorithme est difficile à optimi-

ser. Ainsi, les SVM et les réseaux de neurones apparaissent comme les moins faciles à régler. Néanmoins, comme exposé dans ce chapitre, il est bon de garder à l'esprit que chaque algorithme possède ses propres difficultés d'optimisation même si elles sont plus importantes pour certains. L'optimisation des hyper-paramètres tient un rôle essentiel dans chaque méthode pour obtenir des performances de classification satisfaisantes.

La combinaison de ces deux derniers critères permet d'avoir une meilleure idée de la robustesse globale des algorithmes au surapprentissage. Ainsi, les réseaux de neurones sont généralement les plus sensibles au risque de sur-spécialisation, devant les SVM et les algorithmes basés sur des arbres décisionnels, puis les k-NN.

Toutefois, si ils sont plus laborieux à optimiser, les SVM et les algorithmes fondés sur des réseaux de neurones présentent en contrepartie une meilleure capacité à traiter des problèmes difficiles. Le fonctionnement des réseaux neuronaux avec les fonctions d'activation non-linéaires et cela des SVM avec l'astuce des noyaux (*kernel trick*) leur permettent en effet de modéliser des fonctions non-linéaires plus compliquées que les autres algorithmes.

La complexité d'un problème de classification peut également venir du nombre de classes à traiter. Nous avons vu que tous les algorithmes étudiés sont capable de traiter des problèmes multi-classes. Seul les SVM sont fondés sur un principe de séparation binaire, mais des méthodes existent pour l'adaptation à plus de deux classes en combinant plusieurs SVM simples.

Il est intéressant de constater que cette capacité à résoudre des problèmes difficiles va souvent de pair avec une certaine complexité du fonctionnement. En effet, il est plus difficile de comprendre comment le modèle est construit ou comment la décision de classification est prise pour les SVM et les réseaux de neurones que pour les k-NN, les arbres de décisions et les réseaux bayésiens naïfs dont le fonctionnement est plus intuitif.

Cette interprétabilité de fonctionnement est à ne pas confondre avec l'interprétabilité de la classification, qui caractérise elle la capacité à extraire des informations intelligibles du chemin de classification. Par chemin de classification, nous entendons l'ensemble des étapes qui mène à la décision finale.

Parmi les algorithmes présentés, les arbres décisionnels sont souvent considérés comme très bon pour ce critère. Mais cela n'est pas si facile en pratique car comme expliqué dans ce chapitre, leur interprétabilité se limite premièrement à des arbres de petites tailles. Plus l'arbre est grand, plus les étapes menant à une décision sont nombreuses donc plus il est compliqué d'en extraire une réelle signification. De plus, cette interprétabilité dépend aussi du contexte de l'application donnée. En effet, si les données utilisées pour construire le modèle ont un sens en elles-mêmes (grandeurs physiques par exemple), alors les étapes de classification pourront en avoir également.

Les forêts d'arbres perdent quant à elles en interprétabilité du fait de l'utilisation en parallèle d'un grand nombre d'arbres.

Les réseaux bayésiens naïfs sont des modèles probabilistes qui peuvent fournir une information de degré de certitude pour une classification donnée. De plus, les calculs effectués pour arriver aux probabilités des classes sont simples et intuitifs, ce qui en fait un classifieur présentant également une bonne interprétabilité.

A l'opposé, les algorithmes considérés comme les moins interprétables sont les SVM et réseaux de neurones de part leur fonctionnement comme des boîtes noires. Ils construisent en effet leurs propres représentations des données, ce qui rend le chemin de décision beaucoup moins intelligible.

Pour ce qui est des réseaux de Deep Learning, même si ils sont aussi considérés comme des boîtes noires, de nombreux travaux sont menés à l'heure actuel pour tenter de mieux expliquer leurs prises de décisions, et notamment comment chaque couche de convolution d'un réseau peut se

spécialiser pour réaliser une tâche spécifique. Certaines de ces recherches sont également menées en lien avec les neurosciences pour mieux comprendre le fonctionnement du cerveau humain et développer des systèmes d'intelligence artificielle intelligibles (*Explainable Artificial Intelligence* ou *Transparent Artificial Intelligence* en anglais).

Enfin, pour le cas où les données traitées sont composées d'un grand nombre de dimensions, il est nécessaire de faire attention au phénomène de "malédiction de la dimension" pour la méthode des k-NN. L'algorithme a en effet tendance à être moins efficace avec l'augmentation de la dimensionnalité du problème. Néanmoins, des solutions fondées sur la réduction du nombre de dimensions, comme l'ACP par exemple, existent pour contourner cette difficulté.

Une dimensionnalité des données trop élevée peut également poser problème pour les arbres de décisions, dont le temps de calcul devient alors très élevé pour atteindre l'optimalité. L'utilisation de méthodes d'approximations, ou heuristiques, permet cependant de résoudre ce problème. Les SVM et réseaux de neurones gèrent quant à mieux cette problématique et s'adaptent facilement à des données de grandes dimensions.

4.3.2 Mise en perspective dans le contexte du contrôle d'aspect industriel automatisé

Après avoir mis en avant les avantages et inconvénients des techniques présentées suivant différents critères de comparaison, nous proposons de discuter ceux-ci au regard de leur application au contrôle d'aspect automatisé dans un milieu industriel.

En effet, ce contexte présente des contraintes particulières qui doivent être prises en compte pour choisir un algorithme adapté.

La Figure 4.19 résume graphiquement cette mise en perspective, afin de permettre au lecteur de repérer rapidement les niveaux d'importance des différents critères, détaillés ci-après :

- **Vitesse d'entraînement** : dans un contexte industriel, nous ne considérons pas la vitesse d'entraînement comme critique. En effet, dans notre cas, l'entraînement de l'algorithme se fait en amont de la phase de production en ligne, lors d'une phase de préparation. Le temps d'entraînement peut donc être de l'ordre de quelques minutes à quelques heures sans que cela ne gêne beaucoup le fonctionnement de l'application.
- **Vitesse de classification** : contrairement à la vitesse d'entraînement, la vitesse de classification est critique dans le contexte qui nous intéresse. En effet, le système doit s'adapter aux vitesses de cycles sur des lignes de production, pouvant aller de quelques secondes à quelques dizaines de secondes. Le contrôle doit être réalisé avant la fin du cycle suivant, ce qui implique une classification très rapide ;
- **Occupation mémoire** : même si les dispositifs de stockage actuels offrent des capacités de plus en plus grandes à des coûts de plus en plus faibles, les algorithmes présentant une occupation mémoire faible sont préférés pour une application industrielle ;
- **Robustesse aux données aberrantes** : les données d'apprentissages sont généralement étiquetées par des experts mais le risque de rencontrer des données mal étiquetées existe néanmoins. Il est donc préférable que l'algorithme utilisé soit robuste à la présence de quelques exemples aberrants.
- **Entraînement avec peu d'exemples** : le contexte industriel ne permet en général pas de disposer de beaucoup d'exemple d'apprentissage. En effet, il n'est pas économiquement viable de produire des milliers de pièces pour entraîner un système et il est donc important de pouvoir obtenir des résultats satisfaisants avec un ensemble d'entraînement composé de quelques dizaines voir centaines d'exemples ;

- **Facilité d’optimisation** : la facilitation de réglage de l’algorithme est un critère important dans un contexte industriel, notamment si l’algorithme est destiné à s’adapter à différentes applications ;

- **Complexité des problèmes** : les applications industrielles qui nous intéressent présentent un niveau de complexité relativement variable du point de vue de la classification. Néanmoins, nous considérons que de manière générale, cette complexité reste relativement faible par rapport à d’autres problèmes de classification faisant l’objet de développements à l’heure actuelle comme la reconnaissance de parole, de visages ou encore d’activités dans des vidéos.
En effet, le contrôle de la qualité d’aspect se résume généralement en la classification en deux classes (conforme/non conforme) ou en une classe conforme et plusieurs types de défauts.

- **Interprétation classification** : l’interprétation du chemin de classification peut être utile pour comprendre pourquoi une pièce est déclarée comme non-conforme et ainsi aider à diagnostiquer des dysfonctionnements lors de la production. Néanmoins, ceci est généralement compliqué à mettre en œuvre en pratique car les liens entre les étapes de classification et le processus de production sont souvent complexes ;

- **Interprétation fonctionnement** : la compréhension du fonctionnement de l’algorithme pour mettre au point un modèle ou attribuer une classe à un produit se révèle souvent importante dans le milieu industriel. Ceci est lié à une notion de confiance dans le système de contrôle. En effet, si le fonctionnement de celui-ci est facilement compréhensible, la confiance qui lui sera accordée sera plus grande que pour un système dont le fonctionnement est moins lisible.

- **Dimensionnalité des données** : l’exploitation d’une information riche pour le contrôle visuel tend à engendrer des données avec de grandes dimensions. Cependant, des outils de réduction de la dimensionnalité sont souvent appliqués en amont de la phase de classification (cf. Chap2). Ce critère a donc une importance limitée pour l’algorithme de classification en lui-même si ces outils sont utilisés.

Parmi les critères précédemment exposés, la robustesse au surapprentissage n’a pas été examinée car le contexte considéré ne l’influence pas de manière particulière. En effet un algorithme de classification se veut robuste au phénomène de surapprentissage dans n’importe quel type d’application.

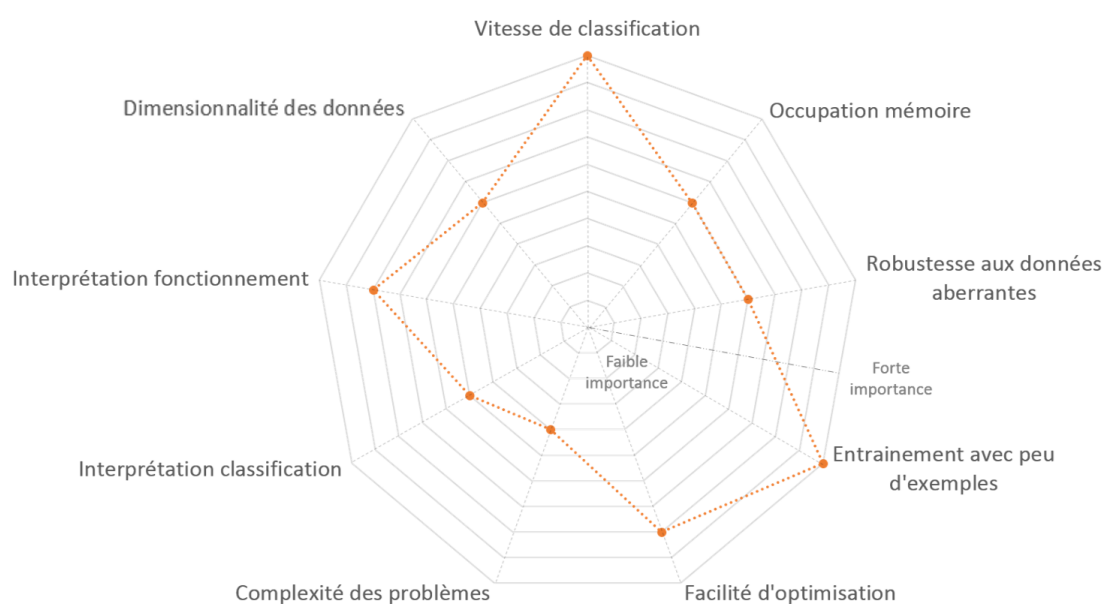


Figure 4.19 – Importance des critères de performances des algorithmes de classification au regard de l'application à l'inspection d'aspect automatisée en industrie

D'après cette mise en perspective et la comparaison présentée précédemment, il apparaît que les algorithmes des k-NN et des réseaux de Deep Learning semblent difficiles à mettre en œuvre dans le contexte visé.

En effet, le désavantage majeur des k-NN est son temps de classification important, rendant difficile son utilisation pour du contrôle en temps réel sur des lignes de production. Néanmoins, dans la pratique les technologies actuelles de calculateurs présentent une puissance de calcul toujours plus élevée, ce qui peut permettre de réduire ce temps de classification, surtout pour des ensembles de données de taille raisonnable.

Pour ce qui est du Deep Learning, les besoins massifs en données d'apprentissage semblent être un obstacle à son utilisation dans un contexte industriel où il est nécessaire de pouvoir entraîner un modèle avec relativement peu d'exemples. Nous souhaitons néanmoins mettre en avant les opportunités qui s'ouvrent à l'heure actuelle avec l'utilisation de réseaux pré-entraînés, qui peuvent permettre de résoudre ce problème grâce notamment à la méthode de *fine tuning partiel*.

En ce qui concerne les autres algorithmes, il est compliqué de prévoir lequel fonctionnera le mieux sur les tâches de classification considérées. Même si il existe des conseils représentés sous formes d'arbres comme celui présenté en Figure 4.20, la question du choix d'un algorithme reste récurrente dans le domaine de l'apprentissage automatique. D'autres exemples sont disponibles en Annexe 1.

Néanmoins, même si de tels mémentos existent, la technique la plus efficace en pratique pour faire un choix reste de calculer la performance de plusieurs algorithmes sur le problème donné, et de choisir celui présentant les meilleurs résultats.

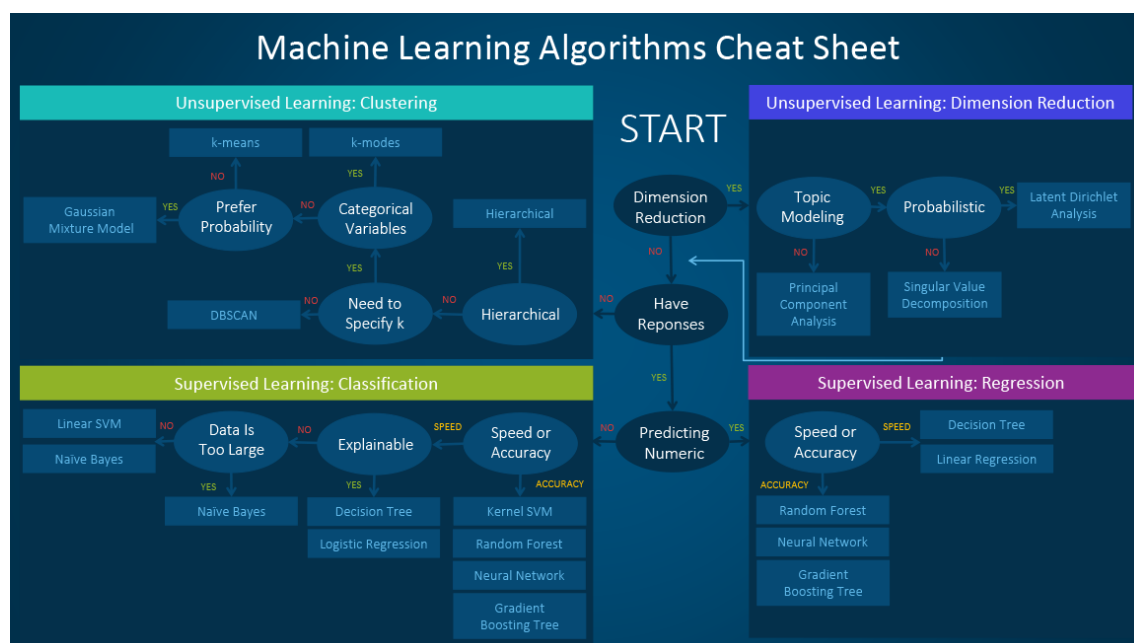


Figure 4.20 – Exemple de cheat sheet pour le choix d’un algorithme d’apprentissage automatisé, issu du [Blog SAS](#)

4.3.3 Méthodes d’ensemble

Au delà du choix d’une technique, il est aussi possible d’utiliser une *combinaison* de plusieurs algorithmes. Ces méthodes combinatoires, aussi appelées *méthodes d’ensemble*, consistent généralement à construire un lot de classifieurs et à classer un nouvel élément grâce à une somme pondérée de leurs prédictions. Dietterich a expliqué en quoi la combinaison de classifieurs peut mener à de meilleures performances que l’utilisation d’algorithme seul [Dietterich2000]. Dans le même article, il a également listé un certain nombre de techniques utilisées dans ce cadre.

Les mécanismes utilisés pour construire des méthodes d’ensemble peuvent être globalement divisés en 3 catégories :

- Construire plusieurs classifieurs du même type mais entraînés sur des sous-ensemble des données d’entraînement. C’est par exemple le cas des forêts d’arbres aléatoires, et plus généralement du procédé de *bagging*, présentées précédemment dans ce chapitre.
- Construire plusieurs classifieurs du même type mais en optimisant différemment les paramètres de chacun d’eux. C’est par exemple le cas de l’utilisation de plusieurs réseaux de neurones en parallèle, en initialisant leurs poids de manière différente, ou en leur donnant des tailles différentes.
- Construire plusieurs classifieurs de différents types. Il est dans ce cas possible d’imaginer de nombreuses combinaisons entre les algorithmes présentés précédemment.

Si ces combinaisons de méthodes ouvrent des perspectives pour l’amélioration des performances de classification, elles ajoutent également de la complexité dans le choix de l’algorithme optimal, de part le grand nombre de combinaisons possibles. Il nous semble néanmoins intéressant de garder ce concept à l’esprit dans la perspective du développement de méthodes génériques.

4.3.4 Intérêt des algorithmes non-supervisés

La comparaison présentée dans cette partie ne prend en compte que des techniques d’apprentissage supervisées. Cependant, nous avons également choisi d’exposer précédemment le fonctionnement de l’algorithme non-supervisé des k-moyennes. Cet algorithme est communément utilisé

4.3. Discussion quant aux caractéristiques des différentes méthodes et leur application industrielle

pour détecter des structures dans des ensembles de données ou plus rarement pour des applications de détection de nouveautés.

Dans le cadre du développement d'un système de contrôle d'aspect automatique, il est important d'être capable de détecter des anomalies déjà connues, mais il peut être aussi très intéressant de détecter des nouvelles anomalies inconnues au moment de l'apprentissage. L'utilisation de techniques non-supervisées comme les k-moyennes pour cette application présente alors un intérêt notable.

4.4 Conclusion générale sur les méthodes de classification

La classification est une étape clé d'un système d'inspection automatisé. Elle permet d'attribuer une classe à un élément à partir de descripteurs extraits auparavant. De nombreuses techniques existent pour réaliser cette tâche, présentant des avantages et des inconvénients les uns par rapport aux autres.

Dans ce chapitre, nous avons tout d'abord introduit le fonctionnement des algorithmes les plus couramment utilisés dans la littérature. Parmi les techniques supervisées classiques, on retrouve les k-NN, les SVM, les arbres et forêts d'arbres décisionnels et les réseaux de neurones.

Chacune de ces méthodes présentent des caractéristiques que nous avons discuté sous forme de comparaison. De manière générale, les SVM et les réseaux de neurones sont capables de traiter des problèmes plus difficiles que les k-NN et arbres décisionnels, mais leur optimisation pour obtenir un modèle satisfaisant est en contrepartie plus compliquée.

Nous avons également proposé d'étudier les caractéristiques de ces algorithmes en regard de leur utilisation dans le contexte du contrôle d'aspect automatisé sur des lignes de production. Il apparaît alors que les k-NN présente un désavantage majeur avec leur temps de classification et leur occupation mémoire importantes. Cependant, nous pensons que cette limite est à relativiser avec le développement de technologies de calculateurs toujours plus rapides et puissants.

Le choix entre telle ou telle technique pour un problème donné est une question récurrente dans le domaine de l'apprentissage automatique dont la réponse n'est pas aisée. Même si il existe des aides pour guider la décision, le moyen le plus sûr reste de tester plusieurs algorithmes pour en déterminer les performances sur le problème donné.

Nous avons également introduit le fonctionnement des réseaux d'apprentissage profond ou *Deep Learning*, car ils connaissent actuellement des développements très importants avec des performances dépassant les techniques d'apprentissage automatique classiques pour certaines tâches. Néanmoins, ils présentent le désavantage majeur d'exiger un lots de données d'apprentissage vaste pour entrainer un modèle valable. Il semble donc difficile de les utiliser pour le moment dans un contexte industriel car il est rarement possible d'obtenir des grands lots de pièces d'apprentissage.

Ensuite, nous avons abordé brièvement le concept des méthodes d'ensemble, qui consistent en la combinaison de plusieurs algorithmes pour améliorer la performance de classification finale. Il semblait intéressant d'aborder ce concept dans le cadre du développement d'un système générique, même si nous n'entrons pas plus dans les détails avec ce mémoire.

Enfin, ce chapitre a également été l'occasion d'introduire l'intérêt que présente les techniques d'apprentissage non-supervisées, comme les k-moyennes par exemple, pour la détection de structures dans un ensemble de données, mais aussi pour la détection de nouveauté. Dans la recherche de généricité, il nous semble en effet important d'être capable de détecter l'apparition de nouvelles anomalies.

Ces quatre premiers chapitres ont permis de tout d'abord situer le contexte de ces travaux de thèse, puis de détailler les différentes étapes du processus d'inspection automatisé, en introduisant certains apports de cette thèse pour les étapes de description d'image et de sélection des données. Pour clore ce mémoire de thèse, nous proposons de présenter dans un dernier chapitre les principes et les résultats expérimentaux du système de contrôle d'aspect automatisé développée au cours de ce travail de thèse. Celui-ci constitue une application reprenant les différentes étapes présentées dans les Chapitre 2 à 4, pour répondre à la problématique d'exploitation d'une information multi-éclairage dans un contexte industriel, énoncées en conclusion du Chapitre 1.

5

Application de l'approche multi-éclairages

Sommaire

5.1	Introduction	164
5.2	Fonctionnement de l'acquisition des images et du traitement des données	165
5.2.1	Acquisition multi-éclairages	165
5.2.2	Algorithme de traitement des données	165
5.2.2.1	Fonctionnements hors-ligne et en-ligne	165
5.2.2.2	Pré-traitement des images	165
5.2.2.3	Description des images	165
5.2.2.4	Algorithme de réduction de l'ensemble des descripteurs	165
5.2.2.5	Algorithmes de classification	165
5.3	Résultats expérimentaux	165
5.3.1	Conditions expérimentales	165
5.3.2	Résultats de classification	165
5.4	Dispositif d'acquisition d'images SPOT	165
5.4.1	Principes de fonctionnement	165
5.4.2	Synthèse sur les degrés de libertés	165
5.5	Conclusion	166

5.1 Introduction

5.2 Fonctionnement de l'acquisition des images et du traitement des données

5.2.1 Acquisition multi-éclairages

5.2.2 Algorithme de traitement des données

5.2.2.1 Fonctionnements hors-ligne et en-ligne

5.2.2.2 Pré-traitement des images

5.2.2.3 Description des images

5.2.2.4 Algorithme de réduction de l'ensemble des descripteurs

5.2.2.5 Algorithmes de classification

5.3 Résultats expérimentaux

5.3.1 Conditions expérimentales

5.3.2 Résultats de classification

5.4 Dispositif d'acquisition d'images SPOT

5.4.1 Principes de fonctionnement

5.4.2 Synthèse sur les degrés de libertés

5.5 Conclusion

Conclusion générale et perspectives

Synthèse des travaux

Dans l'objectif de maintenir un haut niveau de satisfaction chez leurs clients, la maîtrise de la qualité de l'aspect visuel des produits pose un réel défi aux industriels dans de nombreux secteurs d'activités. Or, les processus de contrôle qualité visuel mis en place pour atteindre cette maîtrise se heurtent généralement à une variabilité de résultats due à la subjectivité et à la variabilité du jugement humain.

A partir de ce constat, différents travaux de recherche ont porté sur la mise en place de stratégies pour réduire cette variabilité, soit par la modélisation et la formalisation du contrôle humain pour aider les opérateurs à réaliser le contrôle de manière plus répétable, soit par le développement de systèmes de contrôle automatisés, très répétables par nature.

Ce travail de thèse a porté sur le rapprochement de ces deux approches, en proposant d'exploiter les caractéristiques spécifiques du contrôle humain pour développer un système d'inspection automatisé. Cette idée vient du constat que les dispositifs de contrôle automatique restent encore à l'heure actuelle peu flexibles en matière d'applications. Le développement de systèmes plus génériques en terme de variété de défauts à détecter et de types de pièces à observer constitue donc un challenge pour le domaine de l'inspection automatisée.

Si la répétabilité de l'être humain pour le contrôle d'aspect constitue une faiblesse de ce dernier, sa flexibilité et sa capacité d'adaptation rapide à différents types de produits et d'anomalies en constitue une de ses principales forces. En s'inspirant des pratiques du contrôle visuel humain pour développer un système de contrôle automatisé, et plus particulièrement de l'exploitation de divers angles d'éclairages pour observer les produits, nous souhaitons allier la forte répétabilité des systèmes automatiques avec la flexibilité des pratiques humaines.

Le travail mené dans le cadre de cette thèse a donc permis de montrer qu'il est possible d'exploiter une information multi-éclairages pour mieux détecter des défauts d'aspect, tout en répondant à des contraintes industrielles concernant le temps de traitement.

Pour pouvoir atteindre cet objectif, nous avons utilisé des dispositifs d'acquisition d'images permettant d'acquérir des photos de la pièce contrôlée suivant différents angles d'éclairages. Pour répondre à l'objectif de flexibilité du système, nous avons conçu une nouvelle machine d'acquisition de ce type, qui sera capable d'observer des pièces de grandes dimensions (tenant dans un cube de 50 cm de côté) et de formes variées.

Il a été également nécessaire d'étudier et de mettre en œuvre des moyens permettant de réduire efficacement l'information contenue dans les nombreuses images acquises, pour permettre un temps de traitement en-ligne assez rapide tout en garantissant des performances de classification satisfaisantes.

Enfin, l'étude des différentes méthodes de classification de données a permis de donner des clés pour le choix de méthodes adaptées à ces problématiques de classification dans un contexte industriel.

Les recherches effectuées sur les différentes composantes d'un système d'inspection automatique ont finalement mené au développement d'une application complète de contrôle durant ce travail de thèse. Grâce aux résultats expérimentaux obtenus avec ce dernier, nous avons pu mettre en avant l'intérêt de l'approche multi-éclairages pour la détection de défauts d'aspect.

Au travers de ces développements, nous pouvons mettre en avant les différents apports que nous revendiquons avec cette thèse, mais aussi les différentes perspectives dégagées pour la suite.

Nos apports

Nos apports principaux par ce travail de thèse sont :

- ⇒ le développement d'une méthode permettant d'exploiter une information multi-éclairages pour le contrôle de défauts d'aspect. Nous avons montré l'intérêt de cette approche pour mieux détecter des défauts d'aspect variés et compliqués à observer visuellement.
- ⇒ l'application de la Décomposition Modale Discrète (DMD) à la description d'images dans un objectif de classification automatique.
- ⇒ la proposition d'un nouveau critère de sélection de sous-ensemble de descripteurs fondé sur la statistique du T^2 , pouvant être également utilisé pour faire de la classification. Une étude comparative est néanmoins nécessaire pour pleinement justifier cette approche par rapport aux autres techniques communément utilisées. Elle fait partie des perspectives présentées dans la section suivante.
- ⇒ la conception d'un dispositif de contrôle flexible capable de détecter une grande variété d'anomalies visuelles et de s'adapter à une grande variété de pièces.

Nos apports secondaires par ce travail de thèse sont :

- ⇒ l'application de la DMD pour une application de mesure de performance de la génération de géométries par un réseau adversaire d'apprentissage profond.
- ⇒ l'application de la description d'images à la prédiction de sensations haptiques à partir d'images acquises des surfaces considérées.
- ⇒ une étude comparative des principales méthodes de classification de données communément utilisées au regard de leur application dans un contexte industriel.

Perspectives

Plusieurs perspectives directement reliées aux apports et développements présentés ont tout d'abord été identifiées pour poursuivre nos travaux :

- ⇒ *Concernant le développement de l'approche multi-éclairages :*

Une campagne de mesures est prévue sur un plus grand ensemble de pièces industrielles, avec plusieurs classes de défauts identifiés et annotés par un expert. Cette étude permettra de mettre à l'épreuve l'algorithme développé sur un échantillon plus conséquent, ainsi que de tester sa capacité à discriminer plusieurs classes de défauts plus compliquées à identifier.

- ⇒ *Concernant l'approche modale (DMD) pour la description d'images :*
Il sera intéressant de mener une étude des performances des descripteurs modaux sur différents types d'images afin d'identifier dans quelles situations ils se révèlent les plus pertinents pour la description.
- ⇒ *Concernant le critère de sélection du T^2 :*
Une étude comparative reste à mener pour confronter cette méthode aux autres approches utilisées dans la littérature.
- ⇒ *Concernant l'évaluation de la performance de la génération d'images géométriques avec la DMD :* L'étude du spectre modal à chaque couche du réseau d'apprentissage profond pourrait donner des explications sur le processus de génération des images à travers le réseau.
- ⇒ *Concernant la prédiction de sensations haptiques au travers de la description d'images :*
L'application de ce travail à un lot d'échantillons plus grand permettra de tester de manière plus approfondie la capacité de prédiction du modèle généré. L'extension de ce travail exploratoire à d'autres types de sensations ouvre également de vastes perspectives.
- ⇒ *Concernant la sélection des points d'éclairages :*
Une étude préliminaire a permis de mettre en avant la sensibilité du nombre d'angles d'éclairages nécessaires au type de pièces observées. Le développement d'un algorithme de sélection des points d'éclairages pertinents pour un type de pièces donné, tout en garantissant la flexibilité du système en terme de détection de défauts variés, constitue donc une perspective de ce travail.

D'autres perspectives plus générales sont également proposées pour poursuivre nos travaux :

- ⇒ Dans l'objectif de généricité mis en avant dans notre travail, nous souhaitons également être capables de détecter l'apparition de nouvelles anomalies non prises en compte dans les exemples d'apprentissage. Nous souhaitons donc travailler sur une méthode de classification hybride permettant à la fois de prédire différentes classes de défauts, tout en étant capable d'indiquer si un défaut détecté est inconnu au vu de l'apprentissage.
- ⇒ Nous proposons de tester d'autres types de descripteurs pour décrire les images, notamment issus de réseaux de Deep Learning. Comme introduit dans le chapitre 4, il est en effet possible d'utiliser par exemple certains types de réseaux pré-entraînés pour obtenir des descripteurs pertinents malgré un nombre faible d'exemples d'apprentissage.
- ⇒ Au delà de la détection de défauts d'aspect, une autre problématique d'intérêt liée à l'inspection visuelle des produits réside dans l'évaluation d'écarts de couleurs. Cette problématique constitue notamment un défi important dans l'industrie de l'injection plastique. Il sera intéressant d'étudier l'apport potentiel de l'approche multi-éclairages pour ce type d'application.
- ⇒ Enfin, si l'approche développée au cours de ce travail de thèse a été testée et est destinée pour le moment à l'inspection de produits plastiques, elle pourra également être appliquée au contrôle de produits dans d'autres secteurs comme la métallurgie, l'agroalimentaire ou encore le textile. En effet, notre méthode vise à être générique pour tous types de surfaces contrôlées.

Bibliographie

- [Abouelela2005] Ahmed Abouelela, Hazem M Abbas, Hesham Eldeeb, Abdelmonem A Wahdan et Salwa M Nassar. *Automated vision system for localizing structural defects in textile fabrics*. Pattern recognition letters, vol. 26, no. 10, pages 1435–1443, 2005.
- [Adelson2005] Edward H. Adelson. *Checkershadow Description*. <http://persci.mit.edu/gallery/checkershadow/description>, 2005.
- [Albelwi2017] Saleh Albelwi et Ausif Mahmood. *A framework for designing the architectures of deep convolutional neural networks*. Entropy, vol. 19, no. 6, page 242, 2017.
- [Antonini1992] Marc Antonini, Michel Barlaud, Pierre Mathieu et Ingrid Daubechies. *Image coding using wavelet transform*. IEEE Transactions on image processing, vol. 1, no. 2, pages 205–220, 1992.
- [Anys1995] Hassan Anys et Dong-Chen He. *Evaluation of textural and multipolarization radar features for crop classification*. IEEE Transactions on geoscience and remote sensing, vol. 33, no. 5, page 1170 :1181, 1995.
- [Arthur2007] David Arthur et Sergei Vassilvitskii. *k-means++ : The advantages of careful seeding*. In Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms, pages 1027–1035. Society for Industrial and Applied Mathematics, 2007.
- [Awcock1995] GJ Awcock et R Thomas. *Image Preprocessing*. In Applied Image Processing, pages 90–125. Springer, 1995.
- [Bagot1999] Jean-Didier Bagot. Information, sensation et perception. A. Colin, 1999.
- [Baranowski2012] Piotr Baranowski, Wojciech Mazurek, Joanna Wozniak et Urszula Majewska. *Detection of early bruises in apples using hyperspectral data and thermal imaging*. Journal of Food Engineering, vol. 110, no. 3, pages 345–355, 2012.
- [Baudet2011] Nathalie Baudet, Maurice Pillet et Jean-Luc Maire. *Visual inspection of products : a comparison of the methods used to evaluate surface anomalies*. International Journal of Metrology and Quality Engineering, vol. 2, no. 1, pages 31–38, 2011.
- [Baudet2012] Nathalie Baudet. *Maîtrise de la qualité visuelle des produits-Formalisation du processus d’expertise et proposition d’une approche robuste de contrôle visuel humain*. Thèse de Doctorat, Université de Grenoble, 2012.
- [Bayes1763] Bayes et Price. *An essay towards solving a problem in the doctrine of chances. by the late rev. mr. bayes, frs communicated by mr. price, in a letter to john canton, amfrs*. Philosophical Transactions (1683-1775), pages 370–418, 1763.
- [Baykut2000] Alper Baykut, Alper Atalay, Aytül Erçil et Mustafa Güler. *Real-time defect inspection of textured surfaces*. Real-Time Imaging, vol. 6, no. 1, pages 17–27, 2000.

- [Bennett2000] Kristin P Bennett et Colin Campbell. *Support vector machines : hype or hallelujah ?* Acm Sigkdd Explorations Newsletter, vol. 2, no. 2, pages 1–13, 2000.
- [Blackledge2008] Jonathan Blackledge et Dmitry Dubovitskiy. *A Surface Inspection Machine Vision System that Includes Fractal Texture Analysis*. ISAST TRANSACTIONS ON ELECTRONICS AND SIGNAL PROCESSING, 2008.
- [Blasco2009] J Blasco, N Aleixos, J Gómez-Sanchis et E Moltó. *Recognition and clas-sification of external skin damage in citrus fruits using multispectral data and morphological features*. Biosystems engineering, vol. 103, no. 2, pages 137–145, 2009.
- [Boser1992] Bernhard E Boser, Isabelle M Guyon et Vladimir N Vapnik. *A training algorithm for optimal margin classifiers*. In Proceedings of the fifth annual workshop on Computational learning theory, pages 144–152. ACM, 1992.
- [Bradley1998] Paul S Bradley et Olvi L Mangasarian. *Feature selection via concave mini-mization and support vector machines*. In ICML, volume 98, pages 82–90, 1998.
- [Breiman1984] Leo Breiman. Classification and regression trees. Routledge, 1984.
- [Breiman1996] Leo Breiman. *Bagging predictors*. Machine learning, vol. 24, no. 2, pages 123–140, 1996.
- [Breiman2001] Leo Breiman. *Random forests*. Machine learning, vol. 45, no. 1, pages 5–32, 2001.
- [Briones2006] Vilbett Briones, José M Aguilera et Christopher Brown. *Effect of surface topography on color and gloss of chocolate samples*. Journal of food en-gineering, vol. 77, no. 4, pages 776–783, 2006.
- [Brosnan2004] Tadhg Brosnan et Da-Wen Sun. *Improving quality inspection of food pro-ducts by computer vision :a review*. Journal of Food Engineering, vol. 61, no. 1, pages 3–16, 2004.
- [Bu2009] Hong-gang Bu, Jun Wang et Xiu-bao Huang. *Fabric defect detection based on multiple fractal features and support vector data description*. Enginee-ring Applications of Artificial Intelligence, vol. 22, no. 2, pages 224–235, 2009.
- [Caleb2000] P Caleb et M Steuer. *Classification of surface defects on hot rolled steel using adaptive learning methods*. In KES, pages 103–108, 2000.
- [Campbell1968] Fergus W Campbell et JG Robson. *Application of Fourier analysis to the visibility of gratings*. The Journal of physiology, vol. 197, no. 3, pages 551–566, 1968.
- [Castellini1996] C Castellini, F Francini, G Longobardi, B Tiribilli et P Sansoni. *On-line textile quality control using optical Fourier transforms*. Optics and Lasers in Engineering, vol. 24, no. 1, pages 19–32, 1996.
- [Chan2000] Chi-ho Chan et Grantham KH Pang. *Fabric defect detection by Fourier analysis*. IEEE transactions on Industry Applications, vol. 36, no. 5, pages 1267–1276, 2000.
- [Chang2009] Chuan-Yu Chang, ChunHsi Li, Jia-Wei Chang et MuDer Jeng. *An unsuper-vised neural network approach for automatic semiconductor wafer defect inspection*. Expert Systems with Applications, vol. 36, no. 1, pages 950 – 958, 2009.
- [Chin1982] Roland T Chin. *Automated visual inspection techniques and applications : A bibliography*. Pattern Recognition, vol. 15, no. 4, pages 343–357, 1982.

-
- [Chin1988] Roland T Chin. *Automated visual inspection : 1981 to 1987*. Computer Vision, Graphics, and Image Processing, vol. 41, no. 3, pages 346–381, 1988.
- [Choi2007] Se Ho Choi, Jong Pil Yun, Boyeul Seo, Young Su Park et Sang Woo Kim. *Real-time defects detection algorithm for high-speed steel bar in coil*. In Proceedings of World Academy of Science, Engineering and Technology, volume 21, pages 1307–6884. Citeseer, 2007.
- [Clark Jr2001] John C Clark Jr, Earle W Gillis, Christopher J Majka, Matthew F Seward et Michael M Westgate. *Automated inspection system for metallic surfaces*, March 6 2001. US Patent 6,198,529.
- [CNRS2016] CNRS. Réseaux bayésiens : les formules logiques perdent du poids. <http://www.cnrs.fr/ins2i/spip.php?article2177>, 2016.
- [Cohen1991] Fernand S. Cohen, Zhigang Fan et Stephane Attali. *Automated inspection of textile fabrics using textural models*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 13, no. 8, pages 803–808, 1991.
- [Comon1994] Pierre Comon. *Independent component analysis, a new concept ?* Signal processing, vol. 36, no. 3, pages 287–314, 1994.
- [Conci1998] Aura Conci et Claudia Belmiro Proença. *A fractal image analysis system for fabric inspection based on a box-counting method*. Computer Networks and ISDN Systems, vol. 30, no. 20, pages 1887–1895, 1998.
- [Cortes1995] Corinna Cortes et Vladimir Vapnik. *Support-vector networks*. Machine learning, vol. 20, no. 3, pages 273–297, 1995.
- [Cover1967] Thomas Cover et Peter Hart. *Nearest neighbor pattern classification*. IEEE transactions on information theory, vol. 13, no. 1, pages 21–27, 1967.
- [Cristianini2000] Nello Cristianini et John Shawe-Taylor. *An introduction to support vector machines and other kernel-based learning methods*. Cambridge university press, 2000.
- [Cubero2011] Sergio Cubero, Nuria Aleixos, Enrique Moltó, Juan Gómez-Sanchis et Jose Blasco. *Advances in machine vision applications for automatic inspection and quality evaluation of fruits and vegetables*. Food and Bioprocess Technology, vol. 4, no. 4, pages 487–504, 2011.
- [Dash1997] Manoranjan Dash et Huan Liu. *Feature selection for classification*. Intelligent data analysis, vol. 1, no. 3, pages 131–156, 1997.
- [Debrosse2010] Thomas Debrosse, Maurice Pillet, Jean Luc Maire et Nathalie Baudet. *Sensory perception of surfaces quality-Industrial practices and prospects*. In KEER 2010-INTERNATIONAL CONFERENCE ON KANSEI ENGINEERING AND EMOTION RESEARCH, page 252, 2010.
- [Deeplearning4j2017] Deeplearning4j. Deeplearning4j model zoo. <https://deeplearning4j.org/model-zoo>, 2017.
- [Dennis1989] TJ Dennis et NG Dessipris. *Fractal modelling in image texture analysis*. In IEE Proceedings F-Radar and Signal Processing, volume 136, pages 227–235. IET, 1989.
- [Desage2015] Simon-Frederic Desage. *Contraintes et opportunités pour l’automatisation de l’inspection visuelle au regard du processus humain*. Thèse de Doctorat, Grenoble Alpes, 2015.
- [DGE2016] DGE. *Chiffres clés de l’industrie manufacturière*. Rapport technique, Ministère de l’Économie et des Finances, 2016.
-

- [Diaz2004] R Diaz, L Gil, C Serrano, M Blasco, E Moltó et J Blasco. *Comparison of three algorithms in the classification of table olives by means of computer vision*. Journal of Food Engineering, vol. 61, no. 1, pages 101–107, 2004.
- [Dietterich2000] Thomas G Dietterich. *Ensemble methods in machine learning*. In International workshop on multiple classifier systems, pages 1–15. Springer, 2000.
- [Doak1992] Justin Doak. *An evaluation of feature selection methods and their application to computer security*. Rapport technique, University of California at Davis, 1992.
- [Drucker1996] Harris Drucker et Corinna Cortes. *Boosting decision trees*. In Advances in neural information processing systems, pages 479–485, 1996.
- [ElMasry2012] Gamal ElMasry, Sergio Cubero, Enrique Moltó et José Blasco. *In-line sorting of irregular potatoes by using automated computer-based machine vision system*. Journal of Food Engineering, vol. 112, no. 1, pages 60–68, 2012.
- [Favreliere2007] Hugues Favreliere, Serge Samper, Pierre-Antoine Adragna et Max Giordano. *3D statistical analysis and representation of form error by a modal approach*. In Proceedings of the 10th CIRP Conference on Computer Aided Tolerancing, Specification and Verification for Assemblies, Erlangen Germany, 2007.
- [Favreliere2009] Hugues Favreliere. *Tolérancement modal : de la métrologie vers les spécifications*. Thèse de Doctorat, Université de Savoie, 2009.
- [Fiete1987] RD Fiete et Harrison H Barrett. *Using the Hotelling trace criterion for feature enhancement in image processing*. Optics letters, vol. 12, no. 9, pages 643–645, 1987.
- [Fisher1936] Ronald A Fisher. *The use of multiple measurements in taxonomic problems*. Annals of human genetics, vol. 7, no. 2, pages 179–188, 1936.
- [Forgy1965] Edward W Forgy. *Cluster analysis of multivariate data : efficiency versus interpretability of classifications*. biometrics, vol. 21, pages 768–769, 1965.
- [Freund1996] Yoav Freund, Robert E Schapire et al. *Experiments with a new boosting algorithm*. In Icml, volume 96, pages 148–156. Bari, Italy, 1996.
- [Gilleron2018] Denis & Gilleron. Apprentissage a partir d'exemples - notes de cours. <http://www.grappa.univ-lille3.fr/polys/apprentissage/index.html>, 2018.
- [Goic2011] Gaetan Le Goic, Hugues Favreliere, Serge Samper et Fabien Formosa. *Multi scale modal decomposition of primary form, waviness and roughness of surfaces*. Scanning, vol. 33, no. 5, pages 332–341, 2011.
- [Golnabi2007] H Golnabi et A Asadpour. *Design and application of industrial machine vision systems*. Robotics and Computer-Integrated Manufacturing, vol. 23, no. 6, pages 630–637, 2007.
- [Grandjean2012] Julien Grandjean, Gaëtan Le Goïc, Hugues Favreliere, Yann Ledoux, Serge Samper, Fabien Formosa, Laure Devun et Thomas Gradel. *Multi-scalar analysis of hip implant components using modal decomposition*. Measurement Science and Technology, vol. 23, no. 12, page 125702, 2012.
- [Guerra2008] Anne Sophie Guerra. *Métrologie sensorielle dans le cadre du contrôle qualité visuel*. Thèse de Doctorat, Université de Savoie, 2008.

-
- [Gunasekaran1996] Sundaram Gunasekaran. *Computer vision technology for food quality assurance*. Trends in Food Science & Technology, vol. 7, no. 8, pages 245–256, 1996.
- [Gunning2017] David Gunning. *Explainable artificial intelligence (xai)*. Defense Advanced Research Projects Agency (DARPA), nd Web, 2017.
- [Hajimowlana1998] SH Hajimowlana, R Muscedere, GA Jullien et JW Roberts. *1D autoregressive modeling for defect detection in web inspection systems*. In Circuits and Systems, 1998. Proceedings. 1998 Midwest Symposium on, pages 318–321. IEEE, 1998.
- [Haralick1973] Robert M Haralick, Karthikeyan Shanmugamet al. *Textural features for image classification*. IEEE Transactions on systems, man, and cybernetics, no. 6, pages 610–621, 1973.
- [Hérault1985] Jeanny Hérault, Christian Jutten et Bernard Ans. *Détection de grandeurs primitives dans un message composite par une architecture de calcul neuromimétique en apprentissage non supervisé*. In 10 Colloque sur le traitement du signal et des images, FRA, 1985. GRETSI, Groupe d’Etudes du Traitement du Signal et des Images, 1985.
- [Hinton2003] Geoffrey E Hinton et Sam T Roweis. *Stochastic neighbor embedding*. In Advances in neural information processing systems, pages 857–864, 2003.
- [Holland1975] John H Holland. *Adaptation in natural and artificial systems. An introductory analysis with application to biology, control, and artificial intelligence*. Ann Arbor, MI : University of Michigan Press, pages 439–444, 1975.
- [Hotelling1931] Harold Hotelling. *The Generalization of Student’s Ratio*. Ann. Math. Statist., vol. 2, no. 3, pages 360–378, 08 1931.
- [Hotelling1933] Harold Hotelling. *Analysis of a complex of statistical variables into principal components*. Journal of educational psychology, vol. 24, no. 6, page 417, 1933.
- [Hotelling1951] Harold Hotelling. *A generalized T test and measure of multivariate dispersion*. Rapport technique, UNIVERSITY OF NORTH CAROLINA Chapel Hill United States, 1951.
- [Hu2014] Meng-han Hu, Qing-li Dong, Bao-lin Liu et Pradeep K Malakar. *The Potential of Double K-Means Clustering for Banana Image Segmentation*. Journal of Food Process Engineering, vol. 37, no. 1, pages 10–18, 2014.
- [Huang2015] Szu-Hao Huang et Ying-Cheng Pan. *Automated visual inspection in the semiconductor industry : A survey*. Computers in industry, vol. 66, pages 1–10, 2015.
- [Hung2017] Tzu-Yi Hung, Sriram Vaikundam, Vidhya Natarajan et Liang-Tien Chia. *Phase Fourier Reconstruction for Anomaly Detection on Metal Surface Using Salient Irregularity*. In International Conference on Multimedia Modeling, pages 290–302. Springer, 2017.
- [Iivarinen2000] Jukka Iivarinen. *Surface defect detection with histogram-based texture features*. In Intelligent Systems and Smart Manufacturing, pages 140–145. International Society for Optics and Photonics, 2000.
- [Insana1986] Michael F Insana, Robert F Wagner, Brian S Garra, Reza Momenan et Thomas H Shawker. *Pattern recognition methods for optimizing multivariate tissue signatures in diagnostic ultrasound*. Ultrasonic Imaging, vol. 8, no. 3, pages 165–180, 1986.
-

- [J Mercer1909] BA J Mercer. *XVI. Functions of positive and negative type, and their connection the theory of integral equations.* Phil. Trans. R. Soc. Lond. A, vol. 209, no. 441-458, pages 415-446, 1909.
- [Jeon2009] Yong-Ju Jeon, Jong Pil Yun, Doo-chul Choi et Sang Woo Kim. *Defect detection algorithm for corner cracks in steel billet using discrete wavelet transform.* In ICCAS-SICE, 2009, pages 2769-2773. IEEE, 2009.
- [Jeon2011] Yong-Ju Jeon, Doo-chul Choi, Jong Pil Yun, Changhyun Park et Sang Woo Kim. *Detection of scratch defects on slab surface.* In Control, Automation and Systems (ICCAS), 2011 11th International Conference on, pages 1274-1278. IEEE, 2011.
- [John1994] George H John, Ron Kohavi et Karl Pflieger. *Irrelevant features and the subset selection problem.* In Machine Learning Proceedings 1994, pages 121-129. Elsevier, 1994.
- [Khoje2012] Suchitra Khoje et Shrikant Bodhe. *Performance Comparison of Fourier Transform and its Derivatives as Shape Descriptors for Mango Grading.* International Journal of Computer Applications, vol. 53, no. 3, 2012.
- [Kim2016] Gwan-Hyung Kim, Oh-Hyun Kwon, Sang-il Myong et Sung-Dae Kim. *Development of Plastics Defect Inspection System using Vision System.* International Journal of Control and Automation, vol. 9, no. 7, pages 51-58, 2016.
- [Kiran2016] B Ravi Kiran. K-plus proche voisin (knn/kpp). [https ://beedotkiran.github.io/MLcourse_Lille3/ML_Cours_02_KNN.pdf](https://beedotkiran.github.io/MLcourse_Lille3/ML_Cours_02_KNN.pdf), 2016.
- [Ko2000] Kuk Won Ko et Hyung Suck Cho. *Solder joints inspection using a neural network and fuzzy rule-based classification method.* IEEE Transactions on Electronics Packaging Manufacturing, vol. 23, no. 2, pages 93-103, 2000.
- [Kostal2010] Peter Kostal et Karol Velisek. *Flexible manufacturing system.* World Academy of Science, Engineering and Technology, vol. 53, pages 825-829, 2010.
- [Kotsiantis2007] Sotiris B Kotsiantis, I Zaharakis et P Pintelas. *Supervised machine learning : A review of classification techniques.* Emerging artificial intelligence applications in computer engineering, vol. 160, pages 3-24, 2007.
- [Kudo2000] Mineichi Kudo et Jack Sklansky. *Comparison of algorithms that select features for pattern classifiers.* Pattern recognition, vol. 33, no. 1, pages 25-41, 2000.
- [Kumar2000] Ajay Kumar et Grantham Pang. *Fabric defect segmentation using multi-channel blob detectors.* Optical Engineering, vol. 39, no. 12, pages 3176-3190, 2000.
- [Kumar2002] Ajay Kumar et Grantham KH Pang. *Defect detection in textured materials using optimized filters.* IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), vol. 32, no. 5, pages 553-570, 2002.
- [Kumar2003] Ajay Kumar. *Neural network based detection of local textile defects.* Pattern Recognition, vol. 36, no. 7, pages 1645-1659, 2003.
- [Kumar2008] Ajay Kumar. *Computer-vision-based fabric defect detection : a survey.* IEEE transactions on industrial electronics, vol. 55, no. 1, pages 348-363, 2008.
- [Kuo2003] Chung-Feng Jeffrey Kuo, Ching-Jeng Lee et Cheng-Chih Tsai. *Using a neural network to identify fabric defects in dynamic cloth inspection.* Textile Research Journal, vol. 73, no. 3, pages 238-244, 2003.

-
- [Lacombe2017] Thomas Lacombe, Bruno Albert, Halise Dereli, Mélanie Tomczyk, Youva Nait-Youcef, Jocelyn Philppa et Hughes Favreliere. *Voir pour Toucher ? Caractérisation Optique de la Sensation d'Adhérence*. In QUALITA 2017, 2017.
- [Laplace1820] Pierre Simon Laplace. *Théorie analytique des probabilités*. Courcier, 1820.
- [Latif-Amet2000] A Latif-Amet, Aysin Ertüzün et Aytül Erçil. *An efficient method for texture defect detection : sub-band domain co-occurrence matrices*. Image and Vision computing, vol. 18, no. 6, pages 543–553, 2000.
- [Le Cun1986] Yann Le Cun. *Learning process in an asymmetric threshold network*. In Disordered systems and biological organization, pages 233–240. Springer, 1986.
- [Le Goïc2012] Gaëtan Le Goïc. *Qualité géométrique & aspect des surfaces : approches locales et globales*. Thèse de Doctorat, Université de Grenoble, 2012.
- [LeCun1989] Yann LeCun, Bernhard Boser, John S Denker, Donnie Henderson, Richard E Howard, Wayne Hubbard et Lawrence D Jackel. *Backpropagation applied to handwritten zip code recognition*. Neural computation, vol. 1, no. 4, pages 541–551, 1989.
- [LeCun1990] Yann LeCun, Bernhard E Boser, John S Denker, Donnie Henderson, Richard E Howard, Wayne E Hubbard et Lawrence D Jackel. *Handwritten digit recognition with a back-propagation network*. In Advances in neural information processing systems, pages 396–404, 1990.
- [LeCun1998] Yann LeCun, Léon Bottou, Genevieve B Orr et Klaus-Robert Müller. *Efficient backprop*. In Neural networks : Tricks of the trade, pages 9–50. Springer, 1998.
- [LeCun2015] Yann LeCun, Yoshua Bengio et Geoffrey Hinton. *Deep learning*. nature, vol. 521, no. 7553, page 436, 2015.
- [Leemans2004] Vincent Leemans et M-F Destain. *A real-time grading method of apples based on features extracted from defects*. Journal of Food Engineering, vol. 61, no. 1, pages 83–89, 2004.
- [Li2009a] Te-Sheng Li. *Applying wavelets transform, rough set theory and support vector machine for copper clad laminate defects classification*. Expert systems with Applications, vol. 36, no. 3, pages 5822–5829, 2009.
- [Li2009b] Te Sheng Li et Cheng Lung Huang. *Defect spatial pattern recognition using a hybrid SOM SVM approach in semiconductor manufacturing*. Expert Systems with Applications, vol. 36, no. 1, page 374 : 385, 2009.
- [Li2012] Wu-bin Li, Chang-hou Lu et Jian-chuan Zhang. *A local annular contrast based real-time inspection algorithm for steel bar surface defects*. Applied Surface Science, vol. 258, no. 16, pages 6080–6086, 2012.
- [Li2013] Jiangbo Li, Xiuqin Rao, Fujie Wang, Wei Wu et Yibin Ying. *Automatic detection of common surface defects on oranges using combined lighting transform and image ratio methods*. Postharvest biology and technology, vol. 82, pages 59–69, 2013.
- [Lin2009] Hong-Dar Lin. *Automated defect inspection of light-emitting diode chips using neural network and statistical approaches*. Expert Systems with Applications, vol. 36, no. 1, pages 219–226, 2009.
- [Liu2008] Juan Liu et Hui Ju. *Fuzzy inspection of fabric defects based on particle swarm optimization (PSO)*. In International Conference on Rough Sets and Knowledge Technology, pages 700–706. Springer, 2008.
-

- [Loh1999] Horng-Hai Loh et Ming-Sing Lu. *Printed circuit board inspection using image analysis*. IEEE Transactions on industry applications, vol. 35, no. 2, pages 426–432, 1999.
- [Lu2005] Chi-Jie Lu et Du-Ming Tsai. *Automatic defect inspection for LCDs using singular value decomposition*. The International Journal of Advanced Manufacturing Technology, vol. 25, no. 1-2, pages 53–61, 2005.
- [Lu2008] Chi-Jie Lu et Du-Ming Tsai. *Independent component analysis-based defect detection in patterned liquid crystal display surfaces*. Image and Vision Computing, vol. 26, no. 7, pages 955–970, 2008.
- [Luiz2010] AO Martins Luiz, LC Pádua Flávio et EM Almeida Paulo. *Automatic detection of surface defects on rolled steel using computer vision and artificial neural networks*. In IECON 2010-36th Annual Conference on IEEE Industrial Electronics Society, pages 1081–1086. IEEE, 2010.
- [Maaten2008] Laurens van der Maaten et Geoffrey Hinton. *Visualizing data using t-SNE*. Journal of machine learning research, vol. 9, no. Nov, pages 2579–2605, 2008.
- [MacQueen1967] James MacQueen et al. *Some methods for classification and analysis of multivariate observations*. In Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, volume 1, pages 281–297. Oakland, CA, USA, 1967.
- [Mäenpää2003] Topi Mäenpää, Markus Turtinen et Matti Pietikäinen. *Real-time surface inspection by texture*. Real-Time Imaging, vol. 9, no. 5, pages 289–296, 2003.
- [Mäenpää2004] Topi Mäenpää et Matti Pietikäinen. *Classification with color and texture : jointly or separately ?* Pattern recognition, vol. 37, no. 8, pages 1629–1640, 2004.
- [Mahalanobis1936] Prasanta Chandra Mahalanobis. *On the generalized distance in statistics*. Proceedings of the National Institute of Sciences (Calcutta), vol. 2, pages 49–55, 1936.
- [Maire2013a] J-L Maire, Maurice Pillet et Nathalie Baudet. *Measurement of the perceived quality of a product*. International Journal of Metrology and Quality Engineering, vol. 4, no. 02, pages 63–69, 2013.
- [Maire2013b] Jean-Luc Maire, Maurice Pillet et Nathalie Baudet. *Gage r^2 & e^2 : an effective tool to improve the visual control of products*. International Journal of Quality & Reliability Management, vol. 30, no. 2, pages 161–176, 2013.
- [Majumdar2000] S Majumdar et DS Jayas. *Classification of cereal grains using machine vision : III. Texture models*. Transactions of the ASAE, vol. 43, no. 6, page 1681, 2000.
- [Mak2008] Kai Ling Mak et P Peng. *An automated inspection system for textile fabrics based on Gabor filters*. Robotics and Computer-Integrated Manufacturing, vol. 24, no. 3, pages 359–369, 2008.
- [Mak2009] Kai-Ling Mak, P Peng et KFC Yiu. *Fabric defect detection using morphological filters*. Image and Vision Computing, vol. 27, no. 10, pages 1585–1592, 2009.
- [Malamas2003] Elias N Malamas, Euripides GM Petrakis, Michalis Zervakis, Laurent Petit et Jean-Didier Legat. *A survey on industrial vision systems, applications and tools*. Image and vision computing, vol. 21, no. 2, pages 171–188, 2003.

-
- [Malhi2004] Arnaz Malhi et Robert X Gao. *PCA-based feature selection scheme for machine defect classification*. IEEE Transactions on Instrumentation and Measurement, vol. 53, no. 6, pages 1517–1525, 2004.
- [Mandriota2004] Clelia Mandriota, Massimiliano Nitti, Nicola Ancona, Ettore Stella et Arcangelo Distante. *Filter-based feature selection for rail defect detection*. Machine Vision and Applications, vol. 15, no. 4, pages 179–185, 2004.
- [Mar2011] Nang Seng Siri Mar, PKDV Yarlagadda et C Fookes. *Design and development of automatic visual inspection system for PCB manufacturing*. Robotics and Computer-Integrated Manufacturing, vol. 27, no. 5, pages 949–962, 2011.
- [Masci2012] Jonathan Masci, Ueli Meier, Dan Ciresan, Jürgen Schmidhuber et Gabriel Fricout. *Steel defect classification with max-pooling convolutional neural networks*. In The 2012 International Joint Conference on Neural Networks (IJCNN), pages 1–6. IEEE, 2012.
- [Metropolis1953] Nicholas Metropolis, Arianna W Rosenbluth, Marshall N Rosenbluth, Augusta H Teller et Edward Teller. *Equation of state calculations by fast computing machines*. The journal of chemical physics, vol. 21, no. 6, pages 1087–1092, 1953.
- [Moganti1996] Madhav Moganti, Fikret Ercal, Cihan H Dagli et Shou Tsunekawa. *Automatic PCB inspection algorithms : a survey*. Computer vision and image understanding, vol. 63, no. 2, pages 287–313, 1996.
- [Muller2001] K-R Muller, Sebastian Mika, Gunnar Ratsch, Koji Tsuda et Bernhard Scholkopf. *An introduction to kernel-based learning algorithms*. IEEE transactions on neural networks, vol. 12, no. 2, pages 181–201, 2001.
- [Murthy1998] Sreerama K Murthy. *Automatic construction of decision trees from data : A multi-disciplinary survey*. Data mining and knowledge discovery, vol. 2, no. 4, pages 345–389, 1998.
- [Nagada2006] M Nagada, JG Tallada et T Kobayashi. *Bruise detection using NIR hyper-spectral imaging for strawberry*. Environ. Control Biol, vol. 44, page 133, 2006.
- [Nagorny2018] Pierre Nagorny, Thomas Lacombe, Hugues Favreliere, Maurice Pillet et Eric Pairel. *Generative Adversarial Networks for geometric surfaces prediction in injection molding - Performance analysis with Discrete Modal Decomposition*. In IEEE 19th International Conference on Industriel Technology (ICIT2018), 2018.
- [Narendra1977] Patrenahalli M. Narendra et Keinosuke Fukunaga. *A branch and bound algorithm for feature subset selection*. IEEE Transactions on computers, vol. 9, no. C-26, pages 917–922, 1977.
- [Neogi2014] Nirbhar Neogi, Dusmanta K Mohanta et Pranab K Dutta. *Review of vision-based steel surface inspection systems*. EURASIP Journal on Image and Video Processing, vol. 2014, no. 1, page 1, 2014.
- [Newman1995] Timothy S Newman et Anil K Jain. *A survey of automated visual inspection*. Computer vision and image understanding, vol. 61, no. 2, pages 231–262, 1995.
- [Ng2006] Hui-Fuang Ng. *Automatic thresholding for defect detection*. Pattern recognition letters, vol. 27, no. 14, pages 1644–1649, 2006.
- [Ngan2005] Henry YT Ngan, Grantham KH Pang, SP Yung et Michael K Ng. *Wavelet based methods on patterned fabric defect detection*. Pattern recognition, vol. 38, no. 4, pages 559–576, 2005.
-

- [Ngan2011] Henry YT Ngan, Grantham KH Pang et Nelson HC Yung. *Automated fabric defect detection : review*. Image and Vision Computing, vol. 29, no. 7, pages 442–458, 2011.
- [Nicolai2006] Bart M Nicolaï, Elmi Lötze, Ann Peirs, Nico Scheerlinck et Karen I Theron. *Non-destructive measurement of bitter pit in apple fruit using NIR hyperspectral imaging*. Postharvest biology and technology, vol. 40, no. 1, pages 1–6, 2006.
- [Ojala1996] Timo Ojala, Matti Pietikäinen et David Harwood. *A comparative study of texture measures with classification based on featured distributions*. Pattern recognition, vol. 29, no. 1, pages 51–59, 1996.
- [Pearson1901] Karl Pearson. *LIII. On lines and planes of closest fit to systems of points in space*. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, vol. 2, no. 11, pages 559–572, 1901.
- [Pernkopf2002] Franz Pernkopf et Paul O’Leary. *Visual inspection of machined metallic high-precision surfaces*. EURASIP Journal on Advances in Signal Processing, vol. 2002, no. 7, page 650750, 2002.
- [Pernkopf2003] Franz Pernkopf et Paul O’Leary. *Image acquisition techniques for automatic visual inspection of metallic surfaces*. NDT & E International, vol. 36, no. 8, pages 609–617, 2003.
- [Peterson2007] BS Bennedsen DL Peterson et A Tabb. *Identifying apple surface defects using principal components analysis and artificial neural networks*. Transactions of the ASABE, vol. 50, no. 6, pages 2257–2265, 2007.
- [Pitard2015] Gilles Pitard, Gaëtan Le Goïc, Hugues Favrelière, Serge Samper, Simon-Frédéric Desage et Maurice Pillet. *Discrete Modal Decomposition for surface appearance modelling and rendering*. In Optical Measurement Systems for Industrial Inspection IX, volume 9525, page 952523. International Society for Optics and Photonics, 2015.
- [Pitard2016] Gilles Pitard. *Métrie et modélisation de l’aspect pour l’inspection qualité des surfaces*. Thèse de Doctorat, Grenoble Alpes, 2016.
- [Pitard2017a] Gilles Pitard, Gaëtan Le Goïc, Alamin Mansouri, Hugues Favrelière, Simon-Frédéric Desage, Serge Samper et Maurice Pillet. *Discrete Modal Decomposition : a new approach for the reflectance modeling and rendering of real surfaces*. Machine Vision and Applications, vol. 28, no. 5-6, pages 607–621, 2017.
- [Pitard2017b] Gilles Pitard, Gaëtan Le Goïc, Alamin Mansouri, Hugues Favrelière, Maurice Pillet, Sony George et Jon Yngve Hardeberg. *Robust Anomaly Detection Using Reflectance Transformation Imaging for Surface Quality Inspection*. In Scandinavian Conference on Image Analysis, pages 550–561. Springer, 2017.
- [Porebski2009] Alice Porebski. *Sélection d’attributs de texture couleur pour la classification d’images. Application à l’identification de défauts sur les décors verriers imprimés par sérigraphie*. Thèse de Doctorat, Université Lille 1, 2009.
- [Porwik2015] Piotr Porwik, Rafal Doroz et Tomasz Orczyk. *The k-NN classifier and self-adaptive Hotelling data reduction technique in handwritten signatures recognition*. Pattern Analysis and Applications, vol. 18, no. 4, pages 983–1001, 2015.

-
- [Pottier2014] Thomas Pottier, Hervé Louche, Serge Samper, Hugues Favrelière, Franck Toussaint et Pierre Vacher. *Proposition of a modal filtering method to enhance heat source computation within heterogeneous thermomechanical problems*. International Journal of Engineering Science, vol. 81, pages 163–176, 2014.
- [Pudil1994] Pavel Pudil, Jana Novovičová et Josef Kittler. *Floating search methods in feature selection*. Pattern recognition letters, vol. 15, no. 11, pages 1119–1125, 1994.
- [Quevedo2008] Roberto Quevedo, Fernando Mendoza, JM Aguilera, J Chanona et G Gutierrez-Lopez. *Determination of senescent spotting in banana (Musa cavendish) using fractal texture Fourier image*. Journal of Food Engineering, vol. 84, no. 4, pages 509–515, 2008.
- [Quinlan1993] John Ross Quinlan. C4. 5 : Programs for machine learning. Morgan Kaufmann, 1993.
- [Razmjoooy2012] Navid Razmjoooy, B Somayeh Mousavi et Fazlollah Soleymani. *A real-time mathematical computer method for potato inspection using machine vision*. Computers & Mathematics with Applications, vol. 63, no. 1, pages 268–279, 2012.
- [Rensink1997] Ronald A Rensink, J Kevin O'Regan et James J Clark. *To see or not to see : The need for attention to perceive changes in scenes*. Psychological science, vol. 8, no. 5, pages 368–373, 1997.
- [Rohrmus2005] Dominik R Rohrmus. *Invariant and adaptive geometrical texture features for defect detection and classification*. Pattern Recognition, vol. 38, no. 10, pages 1546–1559, 2005.
- [Rosenblatt1958] Frank Rosenblatt. *The perceptron : a probabilistic model for information storage and organization in the brain*. Psychological review, vol. 65, no. 6, page 386, 1958.
- [Rumelhart1985] David E Rumelhart, Geoffrey E Hinton et Ronald J Williams. *Learning internal representations by error propagation*. Rapport technique, California Univ San Diego La Jolla Inst for Cognitive Science, 1985.
- [Rumelhart1986] David E Rumelhart, Geoffrey E Hinton et Ronald J Williams. *Learning representations by back-propagating errors*. nature, vol. 323, no. 6088, page 533, 1986.
- [Samper2007] Serge Samper et Fabien Formosa. *Form defects tolerancing by natural modes analysis*. Journal of computing and information science in engineering, vol. 7, no. 1, pages 44–51, 2007.
- [Samper2009] Serge Samper, Pierre-Antoine Adragna, Hugues Favreliere et Maurice Pillet. *Modeling of 2D and 3D assemblies taking into account form errors of plane surfaces*. Journal of computing and information science in engineering, vol. 9, no. 4, page 041005, 2009.
- [Schmidhuber2015] Jürgen Schmidhuber. *Deep learning in neural networks : An overview*. Neural networks, vol. 61, pages 85–117, 2015.
- [Shankar2006] NG Shankar et ZW Zhong. *A rule-based computing approach for the segmentation of semiconductor defects*. Microelectronics Journal, vol. 37, no. 6, pages 500–509, 2006.
- [Shannon2001] Claude Elwood Shannon. *A mathematical theory of communication*. ACM SIGMOBILE Mobile Computing and Communications Review, vol. 5, no. 1, pages 3–55, 2001.
-

- [Steinwart2008] Ingo Steinwart et Andreas Christmann. Support vector machines. Springer Science & Business Media, 2008.
- [Stojanovic1999] Radovan D Stojanovic, Stavros A Koubias et George D Papadopoulos. *A Comparison of Statistical and Spectral Analytical Methods for Feature Extraction in the Process of Web Defect Detection*. In IIA/SOCO, 1999.
- [Suvdaa2012] B Suvdaa, J Ahn et J Ko. *Steel surface defects detection and classification using SIFT and voting strategy*. International Journal of Software Engineering and Its Applications, vol. 2, no. 6, pages 161–165, 2012.
- [Thomas1994] ADH Thomas et MG Rodd. *Knowledge-based inspection of electric lamp caps*. Engineering Applications of Artificial Intelligence, vol. 7, no. 1, pages 31–37, 1994.
- [Throop2005] JA Throop, DJ Aneshansley, WC Anger et DL Peterson. *Quality evaluation of apples based on surface defects : development of an automated inspection system*. Postharvest Biology and Technology, vol. 36, no. 3, pages 281–290, 2005.
- [Tibshirani2008] Robert Tibshirani, Jerome Friedman et Trevor Hastie. The elements of statistical learning. Springer, California, 2008.
- [Tien2013] Fang-Chin Tien, Jhih-Syuan Dai, Shih-Ting Wang et Fang-Cheng Tien. *Automated Plastic Cap Defect Inspection Using Machine Vision*. In Proceedings of the Institute of Industrial Engineers Asian Conference 2013, pages 793–800. Springer, 2013.
- [Tolba1997] AS Tolba et AN Abu-Rezeq. *A self-organizing feature map for automated visual inspection of textile products*. Computers in Industry, vol. 32, no. 3, pages 319–333, 1997.
- [Tollari2006] Sabrina Tollari. *Indexation et recherche d’images par fusion d’informations textuelles et visuelles*. Thèse de Doctorat, Toulon, 2006.
- [Tomasi2017] Carlo Tomasi. *Random forest classifiers*, 2017.
- [Tsa2000] D-M Tsa et S-K Wu. *Automated surface inspection using Gabor filters*. The International Journal of Advanced Manufacturing Technology, vol. 16, no. 7, pages 474–482, 2000.
- [Tsai1999] D-M Tsai et C-Y Hsieh. *Automated surface inspection for directional textures*. Image and Vision computing, vol. 18, no. 1, pages 49–62, 1999.
- [Tuceryan1993] Mihran Tuceryan et Anil K Jain. *Texture analysis*. In Handbook of pattern recognition and computer vision, pages 235–276. World Scientific, 1993.
- [Unay2011] Devrim Unay, Bernard Gosselin, Olivier Kleynen, Vincent Leemans, Marie-France Destain et Olivier Debeir. *Automatic grading of Bi-colored apples by multispectral machine vision*. Computers and electronics in agriculture, vol. 75, no. 1, pages 204–212, 2011.
- [Unser1986] Michael Unser. *Sum and difference histograms for texture classification*. IEEE transactions on pattern analysis and machine intelligence, no. 1, pages 118–125, 1986.
- [Vandenbroucke2000] Nicolas Vandenbroucke. *Segmentation d’images couleur par classification de pixels dans les espaces d’attributs colorimétriques adaptés : application à l’analyse d’image*. Thèse de Doctorat, ANRT [diff.], 2000.
- [Vapnik1995] Vladimir N Vapnik. *The Nature of Statistical Learning Theory*. 1995.
- [Vasconcelos2003] Nuno Vasconcelos. *Feature selection by maximum marginal diversity*. In Advances in Neural Information Processing Systems, pages 1375–1382, 2003.

-
- [Verbrugge2004] R Verbrugge, Egon van den Broek, EM van Rikxoort, N Taatgen et L Schomaker. *Evaluation of color representation for texture analysis*. In 16th Belgium-Dutch Conference on Artificial Intelligence, BNAIC 2004, 2004.
- [Wang2014] Zhe Wang et Xiangyang Xue. *Multi-class support vector machine*. In Support Vector Machines Applications, pages 23–48. Springer, 2014.
- [Weil-Barraais1996] A Weil-Barraais. *L'homme cognitif*. PUF, Paris, vol. 576, 1996.
- [Weiwei2008] Liu Weiwei, Yan Yunhui, Li Jun, Zhang Yao et Sun Hongwei. *Automated on-line fast detection for surface defect of steel strip based on multivariate discriminant function*. In Intelligent Information Technology Application, 2008. IITA'08. Second International Symposium on, volume 2, pages 493–497. IEEE, 2008.
- [Wertheimer1923] Max Wertheimer. *A brief introduction to gestalt, identifying key theories and principles*. Psychol Forsch, vol. 4, pages 301–350, 1923.
- [Wettschereck1997] Dietrich Wettschereck, David W Aha et Takao Mohri. *A review and empirical evaluation of feature weighting methods for a class of lazy learning algorithms*. Artificial Intelligence Review, vol. 11, no. 1-5, pages 273–314, 1997.
- [Widrow1960] Bernard Widrow et Marcian E Hoff. *Adaptive switching circuits*. Rapport technique, STANFORD UNIV CA STANFORD ELECTRONICS LABS, 1960.
- [Wilks1932] Samuel S Wilks. *Moments and distributions of estimates of population parameters from fragmentary samples*. The Annals of Mathematical Statistics, vol. 3, no. 3, pages 163–195, 1932.
- [Wiltschi2000] Klaus Wiltschi, Axel Pinz et Tony Lindeberg. *An automatic assessment scheme for steel quality inspection*. Machine Vision and Applications, vol. 12, no. 3, pages 113–128, 2000.
- [Xiao-bo2010] Zou Xiao-bo, Zhao Jie-wen, Li Yanxiao et Mel Holmes. *In-line detection of apple defects using three color cameras system*. Computers and Electronics in Agriculture, vol. 70, no. 1, pages 129–134, 2010.
- [Xie2006] Xianghua Xie, Majid Mirmehdi et Barry Thomas. *Colour tonality inspection using eigenspace features*. Machine Vision and Applications, vol. 16, no. 6, pages 364–373, 2006.
- [Xing2005] Juan Xing, Cedric Bravo, Pal T Jancsok, Herman Ramon et Josse De Baerdemaeker. *Detecting bruises on Golden Delicious apples using hyperspectral imaging with multiple wavebands*. Biosystems Engineering, vol. 90, no. 1, pages 27–36, 2005.
- [Xing2007] Juan Xing, Wouter Saeys et Josse De Baerdemaeker. *Combination of chemometric tools and image processing for bruise detection on apples*. Computers and electronics in agriculture, vol. 56, no. 1, pages 1–13, 2007.
- [Yazdchi2009] Mohammadreza Yazdchi, Mehran Yazdi et Arash Golibagh Mahyari. *Steel surface defect detection using texture segmentation based on multifractal dimension*. In Digital Image Processing, 2009 International Conference on, pages 346–350. IEEE, 2009.
- [Yu2003] Lei Yu et Huan Liu. *Feature selection for high-dimensional data : A fast correlation-based filter solution*. In Proceedings of the 20th international conference on machine learning (ICML-03), pages 856–863, 2003.
-

- [Zhang1995] Yixiang Frank Zhang et Randall R Bresee. *Fabric defect detection and classification using image analysis*. Textile Research Journal, vol. 65, no. 1, pages 1–9, 1995.
- [Zhang2004] Harry Zhang. *The optimality of naive Bayes*. AA, vol. 1, no. 2, page 3, 2004.
- [Zhang2014] Baohua Zhang, Wenqian Huang, Jiangbo Li, Chunjiang Zhao, Shuxiang Fan, Jitao Wu et Chengliang Liu. *Principles, developments and applications of computer vision for external quality inspection of fruits and vegetables : A review*. Food Research International, vol. 62, pages 326–343, 2014.
- [Zhang2016] Hang Zhang, Kristin Dana et Ko Nishino. *Friction from reflectance : Deep reflectance codes for predicting physical surface properties from one-shot in-field reflectance*. In European Conference on Computer Vision, pages 808–824. Springer, 2016.
- [Zhao2010] Jie Zhao, Yongmin Yang et Ge Li. *The cold rolling strip surface defect on-line inspection system based on machine vision*. In Circuits, Communications and System (PACCS), 2010 Second Pacific-Asia Conference on, volume 1, pages 402–405. IEEE, 2010.
- [Zheng2002] Hong Zheng, Ling Xue Kong et Saeid Nahavandi. *Automatic inspection of metallic surface defects using genetic algorithms*. Journal of materials processing technology, vol. 125, pages 427–433, 2002.
- [Zhi2001] Yang Xue Zhi, Grantham KH Pang et Nelson Hon Ching Yung. *Fabric defect detection using adaptive wavelet*. In Acoustics, Speech, and Signal Processing, 2001. Proceedings.(ICASSP'01). 2001 IEEE International Conference on, volume 6, pages 3697–3700. IEEE, 2001.
- [Zhou2014] Wenju Zhou, Minrui Fei, Huiyu Zhou et Kang Li. *A sparse representation based fast detection method for surface defect detection of bottle caps*. Neurocomputing, vol. 123, pages 406–414, 2014.

Table des figures

1	Historique des projet et thèses menées au laboratoire SYMME autour de la problématique de la qualité sensorielle des produits	vi
2	Problématiques liées au développement d'un système d'inspection d'aspect automatisé inspiré du contrôle visuel humain	3
1.1	Problématiques liées à l'acquisition de données inspirée du contrôle visuel humain	8
1.2	Illustration de l'étude d'influences lors du contrôle visuel de surfaces dans un milieu industriel [Debrosse2010]	11
1.3	L'échiquier d'Adelson initial et révélé	12
1.4	Illustion d'optiques inspirées des lois de Gestalt	13
1.5	Illustration de la répétabilité et reproductibilité dans le contexte du contrôle visuel	14
1.7	Modèle pour le contrôle sensoriel proposé par Baudet [Baudet2012]	17
1.8	Les 3 effets lumière et leur caractéristiques [Le Goïc2012]	18
1.9	Exemple de gamme d'exploration proposée par [Le Goïc2012]	18
1.10	Modèle pour le contrôle sensoriel au regard de l'automatisation proposé par Desage [Desage2015]	20
1.11	Parallèle entre le modèle du contrôle sensoriel et une stratégie de vision industrielle selon Desage [Desage2015]	20
1.12	Schéma bloc d'un système typique de vision par ordinateur [Awcock1995]	21
1.13	Parallèle entre le modèle de contrôle sensoriel et une stratégie de vision industrielle plus détaillée	22
1.14	Quelques exemples d'applications pour chaque catégorie de systèmes de vision industrielle selon la classification donnée par Golnabi	25
1.15	Système d'acquisition pour l'inspection de barreaux mettaliques [Li2012]	26
1.16	Système d'acquisition pour le classement de pommes [Leemans2004]	27
1.17	Synthèse des classifications des systèmes de vision industrielle	28
1.18	Points clefs pour le développement d'un système de vision industrielle, dérivés de [Golnabi2007]	30
1.19	Répartition par domaines d'activité des publications référencées un état de l'art récent composé de 125 publications référencées dans le tableau 1.3.2.1	32
1.20	Indicateurs structurels des secteurs industriels en France en 2015 [DGE2016] . .	34
2.1	Problématiques liées à la réduction de l'information par la description d'images .	42
2.2	Illustration d'une représentation du contour et du rectangle englobant minimum pour deux pièces de géométries différentes	44
2.3	Illustration d'une matrice carré aléatoire de 5×5 pixels. Vue numérique (a), représentation en niveaux de gris (b) et visualisation de son histogramme (c). Images issues de [Desage2015].	46
2.4	Illustrations d'images de différentes surfaces rugueuses issues de plaquettes étalons charmillles, de leur transition en niveaux de gris, et enfin de l'histogramme associé. Les surfaces comportent différents niveaux de rugosité (21 = plus petit, 39 = plus grand) et différentes teintes	46

2.5	Illustration du calcul de la matrice de cooccurrences d'une image 5×5 à 4 niveaux de gris, suivant 1 direction ($M1$) puis 4 directions ($M2$) et un pas de 1 dans les deux cas.	47
2.6	Illustrations des matrices de cooccurrences des images précédemment présentées. Afin de mieux mettre en évidence les motifs des matrices, elle sont affichées avec un seuillage sur la colonne de droite.	47
2.7	Illustration du calcul de la nouvelle valeur d'un pixel pour un voisinage direct via la méthodes des LBP [CC : Par Xiawi — Travail personnel, CC BY-SA 3.0]	48
2.8	Trois exemples de voisinages utilisés pour le calcul des LBP [CC : Par Xiawi — Travail personnel, CC BY-SA 3.0]	49
2.9	Illustrations des images précédentes transformées par LBP. A gauche : les images après transformation. A droite : l'histogramme des images transformées.	49
2.10	Illustrations du calcul des histogrammes des sommes et des différences d'une image 3×3 pour un pas de 1 et un angle de 0° puis $0^\circ, 45^\circ, 90^\circ$ et 135°	50
2.11	Exemple de masques associés à différents types de filtres et opérateurs. De gauche à droite : Roberts, Laplacien, Prewitt, Sobel	52
2.12	Illustrations de quelques fenêtres utilisées pour la transformation de Fourier à fenêtre glissante.	53
2.13	Premiers modes de la base modale calculée pour un élément géométrique de référence plan [Le Goïc2012]	56
2.14	Exemple de spectre modal	57
2.15	Représentation de l'analyse multi-échelle d'un spectre modal, issue des travaux de [Le Goïc2012]	59
2.16	Exemple d'analyse multi-échelle sur une mesure de surface métallique, issu de [Le Goïc2012]	60
2.17	Surfaces de réflectance approximées, pour un même pixel par les méthodes PTM, HSH et DMD, issues de [Pitard2016]	61
2.18	Exemple de rendu avec les méthodes PTM, HSH et DMD, pour deux directions d'éclairage S1 et S2, issu de [Pitard2016]	61
2.19	Plaquette charmile présentant des étalons de différentes rugosités	63
2.21	Evolution de certains des paramètres utilisés pour la classification des rugosités. De haut en bas : 14ème paramètre d'Haralick, 1er coefficient modal, maximum et kurtosis de la distribution des coefficients modaux.	67
2.23	Comparaison des spectres modaux des géométries réelles et générées pour deux pièces du lot de test	70
2.24	Résultats des performances de perception de la sensation d'adhérence à l'œil et au toucher	72
2.25	Caractéristiques et illustrations des échantillons utilisés	73
2.26	Comparaison des mesures d'angle de frottement aux résultats de perception haptique	74
2.27	Valeurs d'angles de friction mesurées vs valeurs calculées via le modèle multilinéaire mis au point	75
3.1	Problématiques liées à la réduction de l'information par la sélection de descripteurs	79
3.2	Processus global pour la sélection de paramètres	81
3.3	Arbre de décision partiel d'un problème de sélection <i>branch and bound</i> pour le choix de 3 descripteurs parmi 6	83
3.4	Illustration du fonctionnement d'un algorithme de sélection génétique	87
3.7	Représentation d'un attribut consistant (paramètre 2) et d'un attribut non consistant (paramètre 1) [Porebski2009]	96
3.8	Illustration des axes V_1 et V_2 choisis pour maximiser la variance dans la cadre d'une ACP sur un lot de données décrits au départ par les descripteur D_1 et D_2 . .	100

3.9	Illustration d'une réduction de dimensionnalité via la méthode ACP - (a) Projection des données selon les 3 premières composantes principales - (b) Proportion de la variance expliquée par les composantes principales (vecteurs propres) - (c) Coordonnées des variables initiales dans l'espace de composantes principales (coefficients des combinaisons linéaires)	101
3.10	Illustration du premier axe principal fourni par une ACP et une LDA sur un même ensemble de données	102
3.11	Exemple d'une carte de contrôle du T^2 pour un processus multivarié à 5 variables. 10 observations servent à apprendre les moyennes et la covariance des variables du processus.	107
4.1	Problématiques liées à la classification des données	115
4.2	Principe de fonctionnement de la méthode k-NN pour assigner une classe à une nouvelle donnée, dans un espace de projection composé de 2 descripteurs	118
4.3	Effet du choix de k pour un problème de classification à 2 classes - Haut/Gauche : $k = 1$, Haut/Droite : $k = 2$, Bas/Gauche : $k = 20$, Bas/Droite : $k = 100$ - Exemple issu de [Kiran2016]	119
4.4	Exemple de réseaux bayésien et des tables de probabilités associées [CNRS2016]	121
4.5	Représentation d'un réseau bayésien naïf	122
4.7	Exemple de changement d'espace en passant de coordonnées cartésiennes à des coordonnées polaires	126
4.8	Illustration des <i>variables ressort</i> ξ_i	127
4.10	Principe d'apprentissage d'un arbre de classification via un lot d'entraînement divisé en 3 classes (Figure adaptée de [Tomasi2017]). Les feuilles peuvent être labélisées soit directement avec la classe (rouge, vert ou bleu), soit avec une distribution des probabilités de chaque classe ($p = [p(\text{rouge}), p(\text{vert}), p(\text{bleu})]$). Dans ce cas, toutes les feuilles contiennent une homogénéité maximum car tous les labels présents dans chacune d'elles appartiennent à une même classe.	131
4.11	Structure de boosting, issue de [Drucker1996]	134
4.12	Structure d'un neurone biologique à gauche et artificiel à droite [CC : By Jimmy-neutron - Own work, CC BY-SA 3.0 et Par Chrislb - Made by Chrislb, CC BY-SA 3.0]	135
4.14	Structure d'un perceptron simple, avec N entrées x , une couche cachée de M neurones Z et K les sorties y	137
4.15	Structure d'un perceptron multi couches	141
4.16	Principe et étapes de fonctionnement de l'algorithme des K-moyennes- By Mquantin - Own work, CC BY-SA 4.0	146
4.17	Exemple de deux exécutions sur un même ensemble de données, menant à deux partitionnements différents	146
4.18	Architecture typique d'un réseau de convolution [Albelwi2017]	149
4.19	Importance des critères de performances des algorithmes de classification au regard de l'application à l'inspection d'aspect automatisée en industrie	159
4.20	Exemple de <i>cheat sheet</i> pour le choix d'un algorithme d'apprentissage automatisé, issu du Blog SAS	160
A.1	Fiche de la bibliothèque python <i>Scikit</i>	XXIII
A.2	Fiche de Microsoft Azure Machine Learning	XXIV
A.3	Sphère Mesura : vue éclatée	XXV

Liste des tableaux

1.1	Les 4 descripteurs et leur définition	16
1.2	Tableau répertoriant la bibliographie considérée, composée d'états de l'art, comparatifs et articles présentant des systèmes d'inspection automatisés dans les différents domaines d'applications	33
2.1	Résultat de la classification de 60 pièces de test réparties en 3 classes et des descripteurs d'Haralick	64
2.2	Résultat de la classification de 60 pièces de test réparties en 3 classes et des descripteurs coefficients modaux	65
2.3	Résultat de la classification de 120 pièces de test réparties en 6 classes, avec les descripteurs d'Haralick en bleu, et les coefficients modaux en noir.	65
2.4	Résultat de la classification de 120 pièces de test réparties en 6 classes avec les descripteurs d'Haralick en bleu, et les paramètres statistiques modaux en noir (paramètres statistiques calculés sur le spectre des coefficients)	66
2.5	Similarité entre les spectres	70
3.1	Récapitulatif des algorithmes de génération en fonction de critères situationnels (* : situations rarement rencontrées donc non étudiées pour cette classification - M : Fonction d'évaluation Monotone - NM : Fonction d'évaluation Non Monotone) [Kudo2000] [Porebski2009]	88
3.2	Distances usuelles entre deux points $x = (x_1, x_2, \dots, x_n)$ et $y = (y_1, y_2, \dots, y_n)$ (Σ correspond à la matrice de covariance des données)	93
3.3	Principales caractéristiques des 4 méthodes présentées	104
4.1	Comparaison des algorithmes de classification automatique (++++ très performant / + peu performant)	154



Annexes

A.1 Annexe 1 : Exemples de *cheat sheets* pour la sélection des algorithmes de *machine learning*

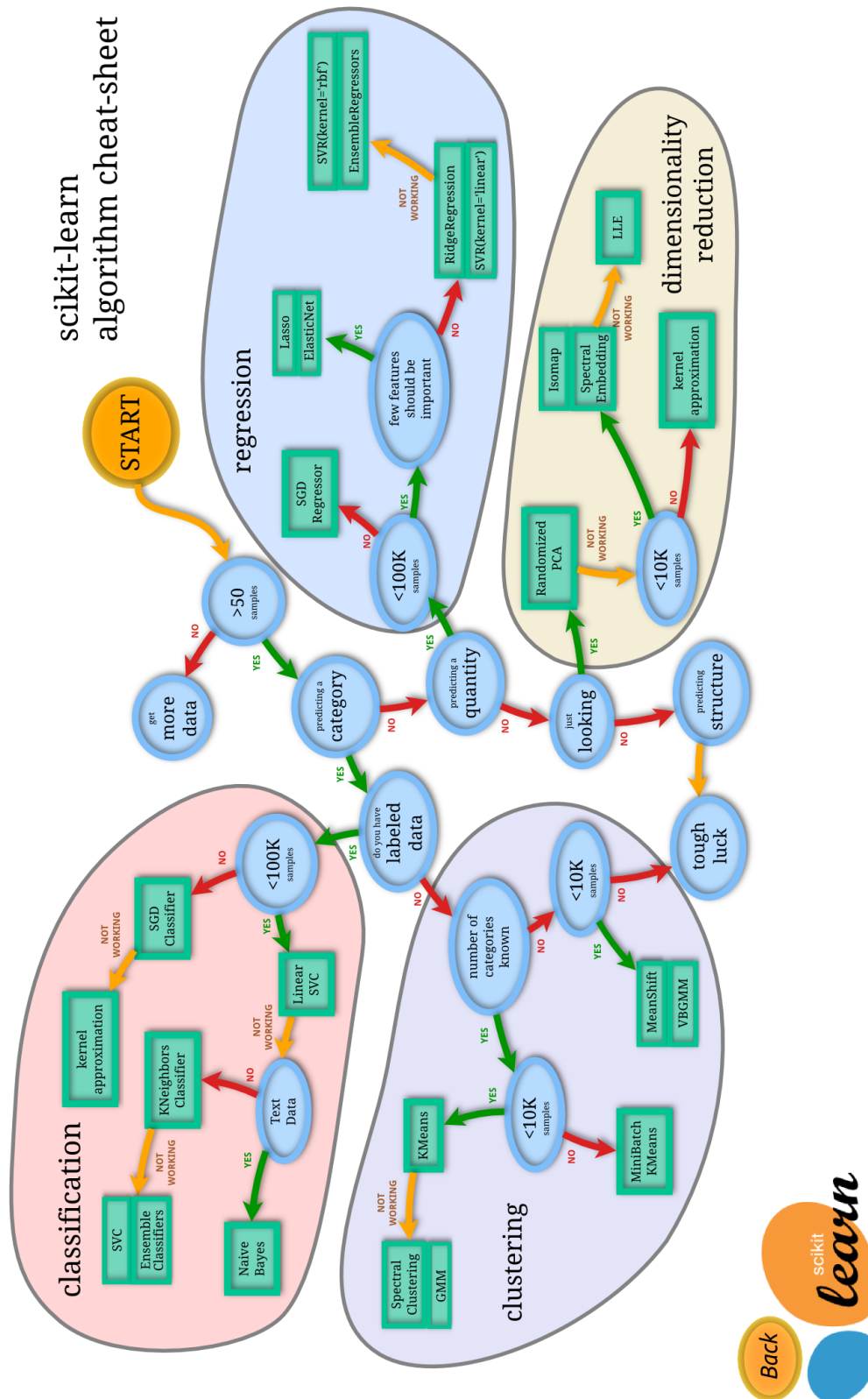


Figure A.1 – Fiche de la bibliothèque python Scikit

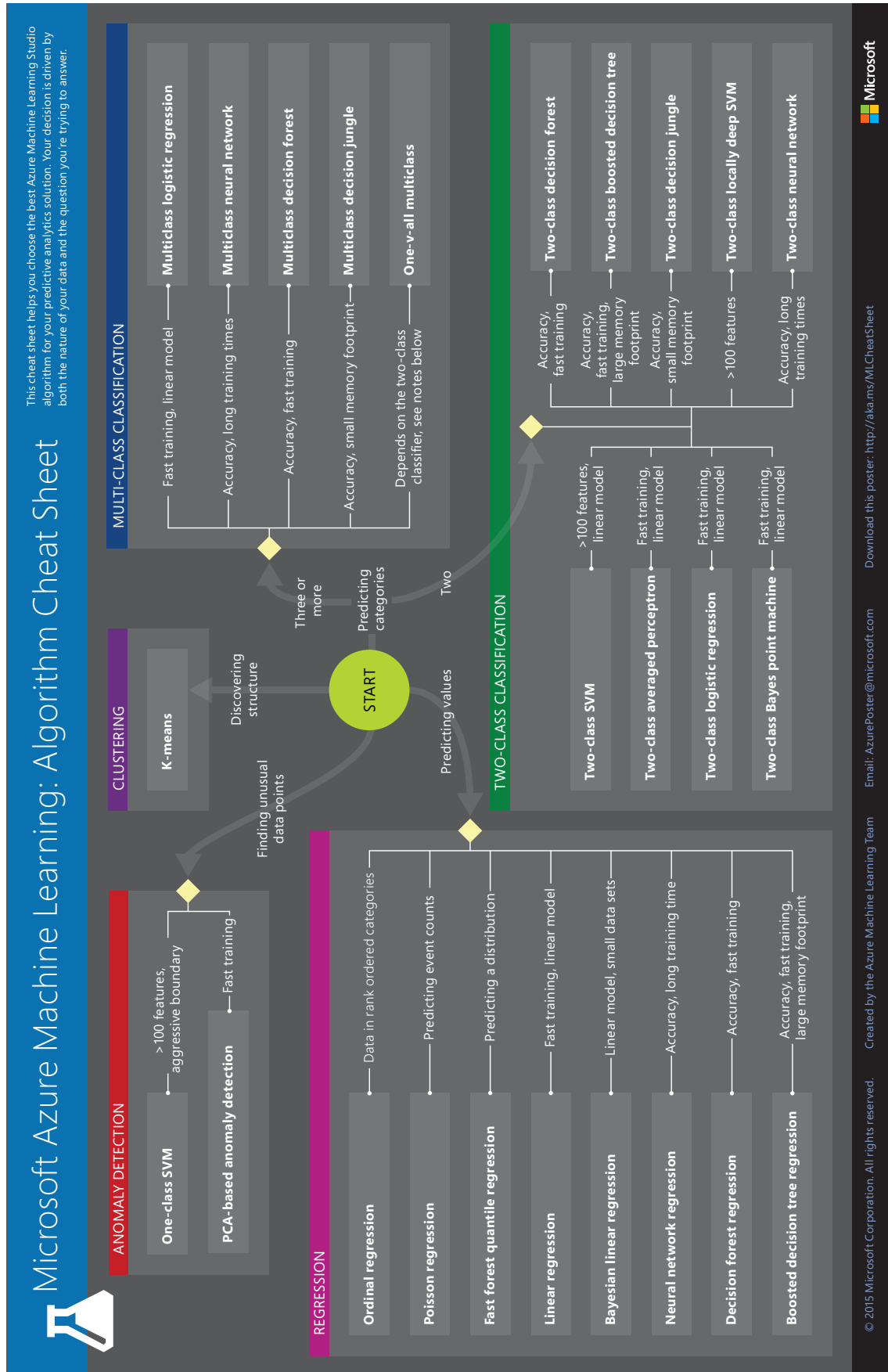


Figure A.2 – Fiche de Microsoft Azure Machine Learning

A.2 Annexe 2 : Représentation détaillée de la sphère Mesura

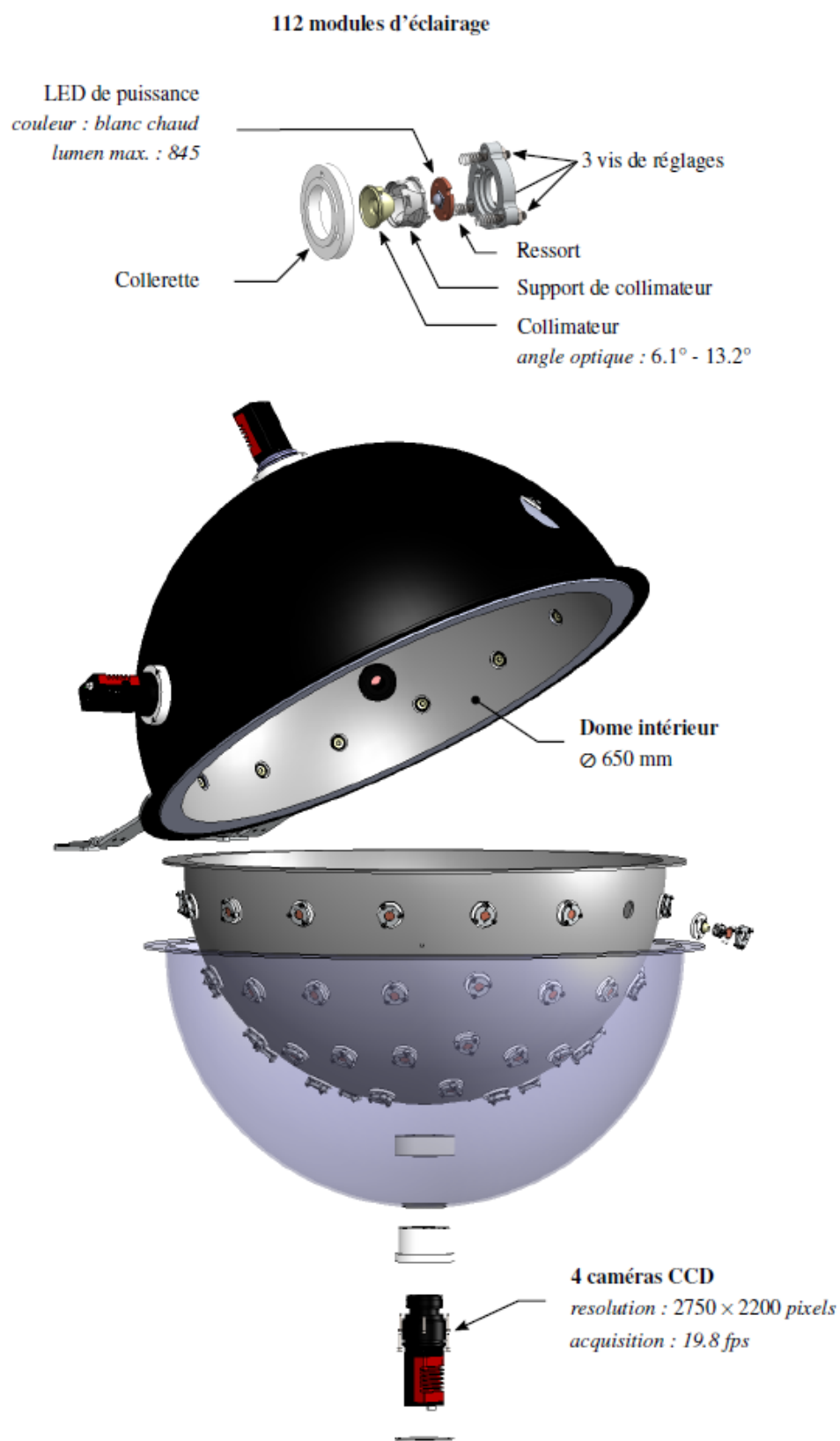


Figure A.3 – Sphère Mesura : vue éclatée

Résumé

La maîtrise de la qualité visuelle des produits constitue un enjeu toujours plus important pour les entreprises à l'heure actuelle, dans de très nombreux secteurs d'application. Afin de garantir cette maîtrise, les industriels se dotent de méthodes et processus axés sur le contrôle visuel des produits. Néanmoins, le contrôle visuel humain en industrie est confronté à des problèmes de variabilité des résultats, dus en partie à l'aspect subjectif de ce type d'inspection. A partir de ce constat, différentes stratégies ont été développées pour réduire cette variabilité. La première consiste en la modélisation et la formalisation du contrôle humain pour aider les opérateurs à réaliser le contrôle de manière plus répétable. La seconde réside dans le développement de systèmes de contrôle automatisés, très répétables par nature, mais cependant encore à l'heure actuelle peu flexibles et adaptables en terme d'applications. Le développement de systèmes plus génériques en terme de variété de défauts à détecter et de types de pièces à observer constitue donc un challenge pour le domaine de l'inspection automatisée.

Ce travail de thèse porte sur le rapprochement de ces deux approches. En s'inspirant des pratiques du contrôle visuel humain pour développer un système de contrôle automatisé, et plus particulièrement de l'exploitation de divers angles d'éclairages pour observer les produits, nous souhaitons allier la forte répétabilité des systèmes automatiques avec la flexibilité des pratiques humaines. Dans cet objectif, une étude détaillée est proposée, partant de la définition du cahier de charges d'un système de contrôle automatisé inspiré du contrôle humain dans un contexte industriel, puis s'intéressant ensuite à ses différentes composantes : l'acquisition de l'information multi-éclairages, l'extraction d'attributs des images et leur réduction, et enfin la classification menant à une décision quant à la qualité du produit inspecté. Pour chaque étape une étude bibliographique des méthodes existantes est menée et mise en regard des contraintes de l'application industrielle. De plus, des contributions méthodologiques innovantes pour l'extraction et la réduction d'attributs sont également proposées. La première consiste en l'utilisation de paramètres issus de la Décomposition Modale Discrète des images comme descripteurs pertinents dans un objectif de classification. La seconde réside dans la proposition d'une nouvelle méthode de sélection d'attributs mettant en œuvre un critère d'évaluation fondé sur la statistique du T^2 et les cartes de contrôle.

L'union des investigations bibliographiques et des contributions méthodologiques issues de ces travaux de thèse mène enfin à la proposition d'une approche innovante exploitant une information multi-éclairages pour le contrôle visuel automatisé. La mise en œuvre de cette approche sur des échantillons issus de l'industrie plastique permet de mettre en avant l'intérêt de cette dernière pour la détection de défaut d'aspect par rapport à une méthode plus classique n'exploitant qu'un seul angle d'éclairage.

Mots-clés : Qualité d'aspect, contrôle qualité industriel, vision industrielle, apprentissage automatique, classification, décomposition modale discrète, imagerie multi-éclairages.

Abstract

Mastering the visual quality of industrial goods is a more and more important concern for companies nowadays, in numerous business sectors. In order to ensure this mastery, industrial companies adopt methods and processes based on visual inspection of products. Nonetheless, human visual inspection in the industry is showing a problematical variability of its results, partly due to the subjective aspect of this kind of control. From this assessment, several strategies have been developed to reduce this variability. The first one consists in modelling and formalizing the human control process to help industrial operators to control more repeatedly. The second lies in developing automated inspection systems, naturally very repeatable, but however still currently not very flexible and adaptable in terms of applications. The development of more generic systems, capable of detecting various types of anomalies on various types of products, is a current challenge for the automated inspection. This thesis work deals with bringing these two approaches together. Drawing inspiration from the human visual control processes to develop an automated control system, and more precisely from the use of various lighting angles to observe a product, we aim to bring together the strong repeatability of automated systems with the flexibility of the human practices. To this end, a detailed study is proposed, from the specifications of an automated inspection system inspired by the human control in an industrial context, to the development of its different components : the acquisition of multi-lighting information, the extraction of features from the images and their reduction, and finally the classification leading to a decision concerning the product's quality. For each step a bibliographic study of the existing methods is produced and contrasted with the industrial constraints. Moreover, some innovative methodological contributions for the extraction and the reduction of features are also proposed. The first one consists in using coefficients coming from Discrete Modal Decomposition of images as relevant features for classification. The second one lies in proposing a new feature selection method using a criteria based on T^2 statistic and control charts. The union between the bibliographic investigations and the methodological contributions coming out of this thesis work leads finally to a proposition for an innovative approach of automated visual control using a multi-lighting information. The implementation of this approach on industrial plastic samples enable to highlight the interest of this method to detect visual defects, in comparison to a more classic method using only one lighting angle.

Keywords : Visual quality, industrial quality control, industrial vision, machine learning, classification, discrete modal decomposition, multi-lighting imaging.