



**HAL**  
open science

# Aide à la détection et à la reconnaissance de défauts structurels dans les pipelines par analyse automatique des images XtraSonic

Clément Fouquet

► **To cite this version:**

Clément Fouquet. Aide à la détection et à la reconnaissance de défauts structurels dans les pipelines par analyse automatique des images XtraSonic. Traitement des images [eess.IV]. Université de Cergy Pontoise, 2014. Français. NNT : 2014CERG0729 . tel-01166341

**HAL Id: tel-01166341**

**<https://theses.hal.science/tel-01166341>**

Submitted on 22 Jun 2015

**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.



Université de Cergy-Pontoise - Ecole doctorale Sciences et Ingénierie

## THÈSE

présentée pour obtenir le titre de DOCTEUR  
en Sciences et Techniques de l'Information et de la communication

---

# AIDE À LA DÉTECTION ET À LA RECONNAISSANCE DE DÉFAUTS STRUCTURELS DANS LES PIPELINES D'HYDROCARBURE À PARTIR D'IMAGES ULTRASONIQUES

---

par

**Clément Fouquet**

ETIS - ENSEA / Université de Cergy-Pontoise / CNRS UMR 8051  
6 avenue du Ponceau, 95014 Cergy-Pontoise Cedex, France

O. ALATA,	Lab. Hubert Curien, Université de St.-Etienne	Examineur
S. BONNET,	TRAPIL, Technopôle de Poissy	Invité
P. DUVAUT,	ETIS, Telecom ParisTech	Directeur de thèse
A. HISTACE,	ETIS, Université de Cergy-Pontoise	Examineur, Co-encadrant
M. PAINDAVOINE,	LEAD, Université de Bourgogne	Rapporteur
F. PRECIOSO,	I3S, Université de Sophia-Antipolis	Rapporteur



## Résumé de la thèse en Français

TRAPIL est une société Française ayant à charge l'exploitation et l'entretien de pipelines d'hydrocarbures. L'entretien de pipelines enterrés nécessite le passage de racleurs équipés de sondes ultrasons réalisant une cartographie de la structure du pipeline, qui est ensuite analysée à la main afin de détecter et d'identifier les différents défauts pouvant apparaître ou évoluer.

L'objectif de ce travail de thèse est d'apporter une solution algorithmique permettant d'accélérer et de compléter le travail des analystes à l'aide des méthodes modernes de traitement d'images et du signal. Notre approche suit le mode opératoire des experts et est découpée en trois parties.

Tout d'abord nous réalisons une détection des soudures d'aboutage permettant de séparer les pipelines en les différents tubes qui le composent. Les signaux de sondes représentant la circonférence du tube sont regroupés et compressés dans une détection de rupture par comparaison de moyenne à court et long terme, puis les signaux résultants sont fusionnés à l'aide d'une pondération unique permettant une augmentation majeure du contraste entre bruit et soudure, offrant une détection et une localisation presque sans faille.

Les tubes subissent ensuite une première segmentation visant à éliminer le plus grand nombre de pixels sains. Usant de modélisation d'histogramme des valeurs d'épaisseur par un algorithme EM initialisé pour notre problématique, l'algorithme suit un principe récursif comparable aux méthodes de type split and merge pour détecter et isoler les zones dangereuses.

Enfin, Les zones dangereuses sont identifiées à l'aide d'une forêt aléatoire, apprise sur un grand nombre d'exemples de défauts. Cette troisième partie est centrée sur l'étude de différentes méthodes de reconnaissance de forme appliquées à notre nouvelle problématique.

Au travers de ces différentes étapes, les solutions que nous avons apportées permettent à TRAPIL un gain de temps significatif sur les tâches les plus fastidieuses du processus d'analyse (par exemple 30% sur la détection de soudures) et leur offre de nouvelles possibilités commerciales, par exemple la possibilité de fournir un pré-rapport à leur clientèle en quelques jours pendant que l'analyse manuelle est réalisée pouvant prendre plus d'un mois.

## Résumé de la thèse en Anglais

TRAPIL is a French society who is in charge of exploitation and maintenance of oil pipelines. Maintenance of buried pipeline implies the use of ultrasonic sensor-equipped devices, providing thickness and structural maps of the pipe, which are analysed by experts in order to detect and identify defects that may appear or evolve. The objective of this work is to provide an algorithmic solution allowing to accelerate and aid the experts's work with modern image and signal processing methods. Our approach follows the experts's operating mode and is divided in three sections.

First, a weld detection is realized allowing to split the pipe in tubes. The signals of probes representing the circumference of the pipe are regrouped and compressed through an abrupt change detection, using short and long-term average comparison, then the resulting signals are merged using a unique weighting function allowing a massive increase of the contrast between welds and noise, offering near-perfect detection and localization.

The tubes then undergoes a first segmentation aiming at eliminating a large amount of sane pixels. Using histogram modelization through an EM algorithm tuned specially for our purpose, the algorithm follows a recursive approach comparable to split and merge methods to detect and isolate dangerous areas.

Finally, those dangerous areas are identified with a Random Forest, which has been learnt on a large amount of defect examples. This third part is greatly focused on the study of different pattern recognition methods applied on our new problematic.

Through those different steps, the solution we brought allows TRAPIL to save a lot of time on the most tedious tasks of the analysis process (for example 30% of gain in processing time for the weld detection) and offers new commercial possibilities, like for example the possibility to provide their clients a first report in a matter of days, while the manual analysis is completed, which can take more than a month.



# Table des matières

<b>Acronymes et notations</b>	<b>ix</b>
<b>Introduction générale</b>	<b>1</b>
Présentation du contexte du travail de thèse : l'entreprise TRAPIL . . . . .	1
La division racleur instrumenté . . . . .	3
Objectifs du travail de thèse et organisation du document . . . . .	3
<b>1 Problématique et objectifs</b>	<b>5</b>
1.1 Matériel utilisé : gamme et fonctionnement des racleurs XTraSonic . . . . .	5
1.2 Problématique de détection de défauts structurels . . . . .	7
1.2.1 Les soudures d'aboutage et les différents types de tubes . . . . .	7
1.2.2 Les différents éléments de ligne . . . . .	8
1.2.3 Les différents types de défauts . . . . .	10
1.2.4 Les erreurs d'acquisition et l'inhibition . . . . .	10
1.2.5 Le processus d'analyse manuelle . . . . .	11
1.3 Objectifs du projet . . . . .	12
1.4 Etat de l'art sur la problématique . . . . .	13
1.4.1 Méthodes utilisées pour ce domaine d'application . . . . .	13
1.4.2 Méthodes utilisées pour des problématiques similaires . . . . .	13
<b>2 Détection des soudures d'aboutage</b>	<b>17</b>
2.1 Introduction . . . . .	17
2.2 Problématique de détection d'événements circonférentiels . . . . .	18
2.3 Méthode proposée : <i>READ-weld</i> . . . . .	19
2.3.1 Vue d'ensemble de <i>READ-weld</i> . . . . .	19
2.3.2 Première étape de <i>READ-weld</i> : Compression et signatures CE . . . . .	19
2.3.3 Seconde étape de <i>READ-weld</i> : Signature unique FCE améliorée . . . . .	20
2.3.4 Troisième étape de <i>READ-weld</i> : Détection et localisation . . . . .	22
2.4 Performances de la méthode <i>READ-weld</i> . . . . .	24
2.4.1 Probabilités de bonnes détections de FCE et probabilités d'erreurs PCE . . . . .	24
2.4.2 Histogrammes de la fonction de détection <i>READ-weld</i> . . . . .	24
2.4.3 Résultats de détection . . . . .	24
2.5 Conclusion . . . . .	25

<b>3</b>	<b>Pré-détection des zones d'intérêt</b>	<b>27</b>
3.1	Introduction . . . . .	27
3.2	Problématique de détection des zones d'anomalie . . . . .	28
3.3	Détection et segmentation des zones d'anomalie ( $DS^2$ ) . . . . .	28
3.3.1	Résumé de la méthode . . . . .	28
3.3.2	Première étape : Prétraitements et détermination de la référence . . . . .	29
3.3.3	Seconde étape : Détection des zones d'anomalie . . . . .	31
3.4	Experimentation et Benchmarking . . . . .	34
3.4.1	Protocole experimental . . . . .	34
3.4.2	Résultats . . . . .	34
3.4.3	Illustration des résultats . . . . .	36
3.4.4	Etude des cas d'échec . . . . .	36
3.5	Conclusion . . . . .	37
<b>4</b>	<b>Identification des zones d'anomalies</b>	<b>39</b>
4.1	Introduction . . . . .	39
4.1.1	Application . . . . .	39
4.1.2	La reconnaissance de forme par apprentissage . . . . .	39
4.1.2.1	Principe général . . . . .	39
4.1.2.2	Le classifieur . . . . .	41
4.1.2.3	Les Forêts aléatoires . . . . .	42
4.2	Problème d'identification des zones d'anomalies . . . . .	43
4.3	Identification des zones d'anomalies . . . . .	44
4.3.1	Aperçu de la méthode . . . . .	44
4.3.2	Première étape : Segmentation fine . . . . .	44
4.3.3	Seconde étape : Extraction de caractéristiques . . . . .	44
4.3.4	Troisième étape : Apprentissage et Classification . . . . .	48
4.4	Performances . . . . .	49
4.5	Conclusion . . . . .	51
	<b>Conclusion générale</b>	<b>55</b>
	Synthèse . . . . .	55
	Déploiement industriel . . . . .	55
	Perspectives . . . . .	56
	<b>Publications</b>	<b>57</b>
	<b>Bibliographie</b>	<b>57</b>

# Table des figures

1	<i>Carte des installations.</i> . . . . .	1
2	<i>Répartition des parts d'action de la compagnie.</i> . . . . .	2
1.1	<i>Exemple de racleur XTraSonic utilisé par TRAPIL. Ici, un XTraSonic 16 pouces 512 sondes</i> . . . . .	5
1.2	<i>Schématisation du tir ultrason pendant le service</i> . . . . .	6
1.3	<i>Exemple de cartographie T2 sur un tube sain, à soudure longitudinale</i> . . .	7
1.4	<i>Exemple de soudure d'aboutage saine. A gauche une photo terrain, à droite une signature T2 acquise par le racleur</i> . . . . .	8
1.5	<i>Exemples de signatures de section de ligne : en haut à gauche un cintre, en haut à droite une gare racleur, en bas à gauche une vanne, et en bas à droite un joint isolant.</i> . . . . .	9
1.6	<i>Exemples de signatures de déviation : de gauche à droite, un té, un split tee, et un piquage.</i> . . . . .	9
1.7	<i>Exemples de signatures de réparation : de gauche à droite, une coquille, un patch, et un manchon.</i> . . . . .	10
1.8	<i>Exemples d'acquisition pour chacune des quatres familles majeures de dé- fauts. A gauche une photo terrain, à droite, une signature obtenue par le racleur sur ce type de défaut. De haut en bas : Enfoncement, Corrosion, Délaminage, Sous-épaisseur.</i> . . . . .	11
1.9	<i>Exemples de tube extrêmement bruité, auquel s'ajoute plusieurs sondes dé- fectueuses (lignes blanches).</i> . . . . .	12
2.1	<i>Portion de ligne utilisée comme exemple dans cette partie. On y voit une signature de soudure au centre et un bruit intense à sa gauche rendant sa détection difficile</i> . . . . .	18
2.2	<i>Schéma méthodologique de READ-weld</i> . . . . .	19
2.3	<i>Signatures CE superposées <math>s_k^{CE}[n]</math>, <math>1 \leq k \leq N_c</math> par rapport à <math>n</math>, avec <math>N_c =</math> <math>28</math>, après traitement de notre exemple Figure 2.1.</i> . . . . .	21
2.4	(a) Signature FCE $s^{FCE}[n]$ par rapport à $n$ , obtenue après combinaison des signatures CE données dans la figure 2.3 au travers de la seconde étape de <i>READ-weld</i> . (b) Contraste $C_{FCE,PCE}^0(s^{FCE})$ de la signature FCE en fonction de l'ordre $L$ des moments $\mu_k^L$ , voir Eq. (2.6) et (2.9). . . . .	23
2.5	<i>Fonction d'ambiguïté <math>D_\lambda^{FCE}[p]</math> basée sur la signature FCE <math>s^{FCE}[n]</math> donnée dans la figure 2.4 (a), après passage dans les sous-étapes 3.1) and 3.2).</i> . . .	23



2.6	(a) Histogramme de $D_{\lambda}^{FCE}[p]$ conditionnellement à $H_{PCE}$ , utilisant <i>READ-weld</i> avec $L = 10$ . (b) Histogramme de $D^*$ conditionnellement à $H_{FCE}$ utilisant <i>READ-weld</i> avec $L = 10$ . Les histogrammes (a) et (b) sont issus du pipeline présenté dans la section 2.4.3 . . . . .	25
2.7	Probabilité de bonne détection $\beta_{FCE}$ vis-à-vis de la probabilité d'erreur $\xi_{PCE}$ pour <i>READ-weld</i> <sub><math>L=10</math></sub> et la méthode REF . . . . .	26
3.1	<i>Portion de tube utilisée à titre d'exemple dans ce chapitre. 256 sondes, pas d'1.5mm. Deux défauts sont présents (cerclés en rouge) : un délaminage à gauche et une légère corrosion à droite. . . . .</i>	28
3.2	<i>Aperçu de la méthode DS<sup>2</sup> . . . . .</i>	29
3.3	<i>Illustration de la découpe en quatre clusters de sondes. . . . .</i>	30
3.4	<i>Résultat de l'algorithme EM sur l'histogramme des valeurs d'épaisseurs de notre exemple illustratif. . . . .</i>	32
3.5	<i>Espace de représentation à deux dimensions (<math>\Delta_{\mu}, \rho_{\sigma}</math>). Les défauts réels sont représentés avec des triangles, les fausses alarmes sont représentés avec des ronds. La séparatrice est en rouge. . . . .</i>	33
3.6	<i>Espace de représentation à deux dimensions (<math>\Delta_{\mu}, \rho_{\sigma}</math>). Les défauts réels sont représentés avec des triangles, les fausses alarmes sont représentés avec des ronds. La séparatrice ad hoc est en rouge, la séparatrice linéaire comparée est en vert. . . . .</i>	35
3.7	<i>Portion de tube utilisée à titre d'exemple dans ce chapitre. 256 sondes, pas d'1.5mm. Deux défauts sont présents (cerclés en rouge) : un délaminage à gauche et une légère corrosion à droite. . . . .</i>	36
3.8	<i>Résultat de la méthode DS<sup>2</sup> sur l'exemple de la figure 3.7. . . . .</i>	36
3.9	<i>Résultat de la méthode par seuillage linéaire sur l'exemple de la figure 3.7. . . . .</i>	37
3.10	<i>Exemples de défauts "perdus" lors de l'experimentation de la méthode DS<sup>2</sup>. Dans le sens de lecture, quatre corrosions, un délaminage, deux enfoncements et une sous-épaisseur. . . . .</i>	38
4.1	<i>Illustration didactique d'un problème de classification où l'on cherche à séparer les images de rose des autres images. . . . .</i>	40
4.2	<i>Nuage de points sur la base de donnée Iris avec deux caractéristiques : longueur du sépale et la longueur des pétales. . . . .</i>	41
4.3	<i>Exemple de déroulement de l'algorithme d'un arbre de décision binaire sur un cas simple à deux classes et deux caractéristiques . . . . .</i>	43
4.4	<i>Exemples de résultats de binarisation sur des extractions de défauts réalisées au chapitre précédent. . . . .</i>	45
4.5	<i>Pourcentages de bonne détection pour chaque classes de défauts avec rééquilibrage de la base par suppression de fausses alarmes. . . . .</i>	52
4.6	<i>Pourcentages de bonne détection pour chaque classes de défauts avec rééquilibrage de la base par augmentation artificielle du nombre de défauts. . . . .</i>	53

# Liste des tableaux

1.1	<i>Table de correspondance entre le diamètre de la ligne, nombre de sonde du racleur et pas d'acquisition</i> . . . . .	6
1.2	<i>Liste des types de défauts caractérisés par le POF</i> . . . . .	15
3.1	<i>Performances des méthodes comparées sur la ligne de test</i> . . . . .	35
4.1	<i>Propriétés de symétries des matrices de co-occurrence</i> . . . . .	46
4.2	<i>Matrice de confusion résultat pour le PMC sur la base de tests avec suppression de fausses alarmes</i> . . . . .	49
4.3	<i>Matrice de confusion résultat pour le PMC sur la base de tests avec augmentation artificielle du nombre de défauts</i> . . . . .	50
4.4	<i>Matrice de confusion résultat pour le SVM sur la base de tests avec suppression de fausses alarmes</i> . . . . .	50
4.5	<i>Matrice de confusion résultat pour le SVM sur la base de tests avec augmentation artificielle du nombre de défauts</i> . . . . .	50
4.6	<i>Matrice de confusion résultat pour la forêt aléatoire sur la base de tests avec suppression de fausses alarmes</i> . . . . .	50
4.7	<i>Matrice de confusion résultat pour la forêt aléatoire sur la base de tests avec augmentation artificielle du nombre de défauts</i> . . . . .	51



# Acronymes et notations

## Acronymes

- *ACP* (Analyse en Composante Principale) : Méthode de fouille de donnée permettant de déterminer la corrélation entre les caractéristiques d'un jeu de données.
- *Bagging* (Bootstrap Aggregating) : Méthode de classification qui consiste à créer des sous ensemble d'échantillons tirés aléatoirement avec remise.
- *CE* (Circumferential Event) : Evénement circonférentiel.
- *COFRAC* (COmité FRançais d'ACcréditation) : Instance française d'accréditation.
- *dB* (décibel) : Unité de mesure acoustique utilisée pour caractériser le bruit ou un contraste entre deux signaux.
- *DS<sup>2</sup>* (Detection and Segmentation of Defect Signature) : Nom de la méthode proposée de pré-détection des zones d'intérêt.
- *EM* (Expectation - Maximization) : Algorithme permettant la modélisation d'histogrammes.
- *FCE* (Full Circumferential Event) : Evénement se répercutant sur la totalité de la circonférence du pipeline.
- *GLCM* (Gray Level Co-occurrence Matrix) : Matrice de co-occurrence.
- *GLD* (Gottsborg Leak Detector) : Détecteur de fuite par mesure d'acoustique à l'aide d'un hydrophone.
- *LO* : Méthode d'acquisition par sonde ultrason où la sonde est positionnée perpendiculairement à la pièce de métal à analyser.
- *LARA* (Logiciel d'Analyse RAcleur) : Logiciel utilisé par les analyste de TRAPIL pour lire les fichiers de données racleur.
- *LHP* (Le Havre - Paris) : Réseau de pipeline reliant le port du Havre à Paris.
- *MATLAB* (MATrix LABoratory) : Logiciel contenant un grand nombre de fonctions et de *toolbox* utiles au développement de méthodes de traitement du signal.
- *MFL* (Magnetic Flux Leakage) : Détecteur de défauts par analyse de flux magnétiques envoyés dans le métal.
- *OCR* (Optical Character Recognition) : Logiciel de reconnaissance de caractères manuscrits ou imprimés.
- *ODC* (Oléoduc de Défense Commune) : Pipeline appartenant à l'OTAN permettant le ravitaillement des bases militaires.
- *PA* (Position Angulaire) : Index correspondant à l'angle formé entre la première sonde du racleur et la sonde ayant effectué le tir.
- *PCE* (Partial Circumferential Event) : Evénement ne se répercutant que partiellement sur la circonférence du pipeline.
- *PK* (Point Kilométrique) : Index correspondant à la distance parcourue par le

- racleur, depuis son départ, au moment du tir.
- *PMC* (Perceptron Multi-Couches) : Autre nom du réseau de neurone, un classifieur.
  - *PMR* (Pipeline Méditerranée - Rhône) : Pipeline situé dans le sud-est de la France.
  - *POF* (Pipeline Operators Forum) : Forum composé des différents exploitants de pipeline en Europe.
  - *READ-weld* (REcognition and Automated Detection of welds) : Nom de la méthode proposée de détection des soudures d'aboutage.
  - *RF* (Random Forest) : Forêt aléatoire, un classifieur.
  - *RFS* (Random Feature Selection) : Méthode permettant d'injecter de l'aléatoire dans les arbres de décision binaire.
  - *SVM* (Support Vector Machine) : Machine à vecteur de support, un classifieur.
  - *WEKA* : Logiciel permettant le test de classifieurs sur des bases de données formatées.
  - *ZA* (Zone d'Anomalie) : Zone définie comme pouvant contenir un défaut.
  - *ZAT* (Zone Affectée Thermiquement) : Zone entourant une soudure d'aboutage où le métal a été chauffé et donc fragilisé.

## Notations

### Chapitre 1

- $N$  : Nombre de pixels d'une ligne entière.
- $S$  : Nombre de sondes du racleur.
- $L$  : Longueur de la ligne entière.
- $k$  : Pas d'acquisition utilisé par le racleur.
- $A1$  : Source de donnée représentant l'amplitude du signal lors du contact avec la paroi interne.
- $A2$  : Source de donnée représentant l'amplitude du signal lors du contact avec la paroi externe.
- $T1$  : Source de donnée représentant le temps de parcours racleur - paroi interne.
- $T2$  : Source de donnée représentant le temps de parcours paroi interne - paroi externe.

### Chapitre 2

- $H$  : Hypothèse d'appartenance du pixel considéré.
- $C_{a,b}^\pi(\cdot)$  : Opérateur de contraste entre deux hypothèses.
- $s[n]$  : Signature distribuée.
- $N_c$  : Nombre de clusters.
- $N_s$  : Nombre de sondes.
- $q$  : Nombre de sondes par cluster.
- $\mu_{k,short}[n]$  : Moyenne mouvante des valeurs situées en amont du pixel concerné.
- $\mu_{k,long}[n]$  : Moyenne mouvante des valeurs situées en aval du pixel concerné.
- $\alpha_k[n]$  : Fonction de pondération.
- $\mu_k^L[n]$  : Moment d'ordre  $L$  de la signature  $s_k^{CE}[n]$ .
- $L$  : Ordre du moment utilisé.
- $\lambda$  : Valeur d'échelle.
- $p$  : Index de localisation.

- $D_{\lambda}^{FCE}[p]$  : Fonction d'ambiguïté réalisée en projetant le signal sur la base d'ondelettes de Haar.
- $\beta$  : Probabilité de bonne détection.
- $\eta$  : Probabilité d'erreur.

### Chapitre 3

- $H$  : Hypothèse d'appartenance du pixel considéré.
- $L$  : niveau de recursivité courant.
- $N_c$  : Nombre de clusters.
- $K$  : Nombre de noyaux utilisés pour l'initialisation de l'EM.
- $P$  : Probabilité d'appartenance.
- $\mu$  : Moyenne.
- $\sigma$  : Ecart-type.
- $\Delta_{\mu}$  : perte de métal relative.
- $\rho_{\sigma}$  : diversité des valeurs d'épaisseur présentes.
- $\%D$  : Pourcentage de défauts réels présents dans les zones retournées par rapport au nombre de défauts d'origine.
- $nbZA$  : Nombre de zones d'anomalies retournées.
- $\%P$  : Pourcentage de pixels retournés par rapport au nombre de pixels d'origine.

### Chapitre 4

- $\gamma$  : Paramètre utilisé dans la fonction de base radiale,  $\gamma = -\frac{1}{2\sigma^2}$ .
- $\sigma$  : Hyperparamètre du SVM.
- $t$  : Translation.
- $d$  : Distance de la translation  $t$ .
- $\theta$  : Angle formé entre l'horizontale et la translation  $t$ .
- $N$  : Nombre de niveaux de gris utilisés pour coder l'image source.
- $K$  : Nombre de translations de niveaux de gris dans l'image originale.



# Introduction générale

## Présentation du contexte du travail de thèse : l'entreprise TRAPIL

TRAPIL est une entreprise française qui exploite des réseaux de pipeline pour le transport de produits pétroliers. Elle est créée en 1950 suite à la loi du 2 août 1949 avec pour objet la construction et l'exploitation du réseau d'oléoducs Le Havre - Paris (LHP).

En 1953, l'entreprise met en service une première canalisation de 200km de long reliant les raffineries de Normandie aux dépôts parisiens. Le LHP démarre alors. Par la suite TRAPIL étend son réseau au point de doubler dans les années 1960 et de finir par tripler. En parallèle de l'extension de son réseau, TRAPIL se voit attribuer l'exploitation des Oléoducs de Défense Commune de l'OTAN (ODC) et du Pipeline Méditerranée - Rhône (PMR)

Propriétaire de son réseau civil et prestataire des ODC et du PMR, TRAPIL opère aujourd'hui sur près de 4700km de canalisation, 160 installations de pompage et livraison, et 850 000  $m^3$  de stockage. Toutes ces installations s'étendent sur une grande partie du territoire français (voir Fig. 1). Le siège social de la société est situé à Paris dans le XV<sup>e</sup> arrondissement aux 7 et 9 rue des frères Morane.

L'entreprise initialement publique est devenue totalement privée avec la vente des parts de l'état français. Son capital est aujourd'hui réparti entre de grandes sociétés pétrolières et ses salariés. Parmi ces sociétés, Total est l'actionnaire majoritaire (voir Fig. 2). A l'issue de l'année 2012, TRAPIL disposait d'un capital de 13 240 800 € et a enregistré un chiffre

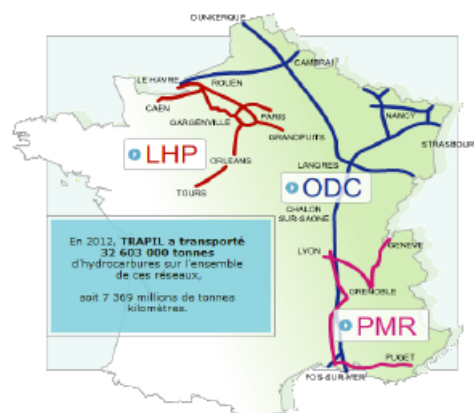


FIGURE 1 – Carte des installations.



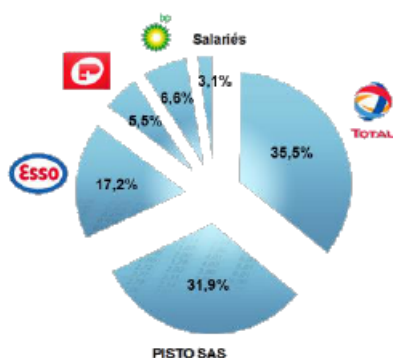


FIGURE 2 – Répartition des parts d'action de la compagnie.

d'affaires de 154 094 000 € dont 26 711 412 € de bénéfices.

Afin d'assurer l'exploitation permanente de ces réseaux, TRAPIL emploie plus de 750 salariés répartis sous 6 directions :

- Réseaux pipelines civils
- Réseau pipeline OTAN
- Technique
- Administrative et financière
- Ressources humaines
- Qualité, hygiène, sécurité et environnement

Bien que son activité principale soit le transport de produits pétroliers, l'entreprise propose aussi une gamme de prestations dérivées de son expérience et agit dans quatre autres domaines d'activités :

- Ingénierie et conseils : Le conseil en exploitation peut s'appliquer à des points particuliers de l'exploitation ou concerner une réflexion approfondie sur les méthodes et les moyens à mettre en oeuvre pour apporter une réponse à des questions d'exploitation. Les prestations d'ordre techniques et / ou administratives, couvrent l'ingénierie des pipelines et l'ingénierie des dépôts.
- Planification et automatisation : TRAPIL conçoit, adapte et met en oeuvre de nouveaux moyens informatiques dans les domaines industriels et scientifiques liés à l'exploitation des pipelines tels que les automatismes complexes des installations, les calculs des plans de pompage, le comptage, etc. En matière de planification, ces outils informatiques permettent à une équipe d'élaborer les programmes de transport en intégrant l'ensemble des contraintes liés à l'exploitation d'un réseau de pipelines multiproduits.
- Etalonnage et analyses : TRAPIL bénéficie d'un savoir-faire très large en matière d'étalonnage et de métrologie assuré par des laboratoires accrédités par le COFRAC (COmité FRançais d'ACcréditation) dans les domaines de la débitmétrie, de la masse volumique, de la température et de la pression. L'entreprise dispose également d'un laboratoire d'analyses accrédité COFRAC dédié aux analyses d'hydrocarbures.
- Inspection par racleurs instrumentés : Dans le domaine de la mesure et de l'instrumentation, TRAPIL a mis au point des outils de contrôle intra-tubes : les racleurs instrumentés. Ils ont pour but de contrôler l'intégrité des tubes, ainsi que les défauts dus à la corrosion ou à la fissuration.

C'est au sein de la division racleur instrumenté que prend place ce travail de thèse.

## La division racleur instrumenté

Le pipeline est un outil essentiel au transport massif de fluide : discret, non polluant et possédant un débit important de transport. Cependant leur enfouissement ne rend pas leur maintenance aisée. Des outils permettant l'inspection du pipeline depuis l'intérieur ont été développés : les racleurs instrumentés. L'idée est d'utiliser les racleurs, originellement des gabarits équipés de coupelles qui étaient envoyés dans le tube afin de le nettoyer, en leur intégrant un système mécanique ou électronique permettant de vérifier l'intégrité de la canalisation. TRAPIL dispose de trois types de racleurs instrumentés :

- Le contrôle d'étanchéité mesure l'acoustique par hydrophone. Il enregistre les variations dans une bande de fréquence précise et permet de localiser une fuite.
- Le calibreur effectue un contrôle de géométrie. Il est doté de palpeurs mécaniques sur toute sa circonférence et mesure ainsi les variations de rayon interne. Il est capable de détecter les soudures d'aboutage et les déformations du tube de grande ampleur. Il est souvent utilisé avec une plaque gabarit en amont d'un passage XTraSonic.
- L'XTraSonic enfin, effectue un contrôle double de géométrie et de manque de métal. Il est doté de sondes ultrasons sur sa circonférence. Par les mécanismes physiques de réflexion et de réfraction, l'outil permet de mesurer le rayon interne mais aussi l'épaisseur du tube.

Un passage racleur XTraSonic se déroule comme suit :

- Le racleur est préparé et vérifié en atelier selon les spécifications de la ligne à analyser.
- Le racleur est ensuite acheminé par camion à la gare racleur de départ, où il sera installé, testé et mis en service avant d'être lancé dans le pipeline.
- Le racleur est mu par le débit du fluide transporté par le pipeline et enregistre les données de manière autonome.
- Une fois le passage réalisé, le racleur est récupéré à la gare d'arrivée et ramené dans les locaux de TRAPIL, où les données acquises sont extraites et transmises à la section analyse.
- La section analyse formate les données, et réalise l'analyse complète du passage, confectionnant un rapport d'analyse contenant toutes les informations utiles et l'envoi au client.

Ce travail de recherche de type R&D se positionne à la dernière étape, lors de l'analyse du passage.

## Objectifs du travail de thèse et organisation du document

L'objectif du travail de thèse se situe au niveau de l'analyse, en bout de chaîne du processus d'inspection par racleur instrumenté. En effet le travail d'analyste consiste à passer en revue une ligne de pipeline entière, pouvant mesurer des centaines de kilomètres, centimètre par centimètre ; c'est un travail fastidieux et long, pouvant nécessiter plusieurs mois/homme. L'objectif initial de ce travail de thèse est donc de proposer, à l'aide des méthodes modernes de traitement d'image et du signal, une solution logicielle d'automatisation permettant d'alléger et d'accélérer le travail des analystes.

Le document est organisé comme suit :

- Tout d’abord la problématique est posée de manière plus technique avec les détails et difficultés du travail d’analyste.
- Nous présentons par la suite une solution apportée à une première problématique fondamentale : la détection des soudures d’aboutage.
- Puis nous proposons une méthode permettant de réaliser une première segmentation des défauts à détecter.
- Ensuite nous présentons le travail de recherche réalisé sur l’application de méthodes de reconnaissance de formes afin d’identifier les défauts pré-segmentés.
- Enfin nous concluons, en rappelant les résultats obtenus, en expliquant comment le travail accompli permet d’apporter une aide réelle quantifiée au travail de l’analyste et en présentant les principales perspectives de ce projet.

# Chapitre 1

## Problématique et objectifs

### 1.1 Matériel utilisé : gamme et fonctionnement des racleurs XTraSonic

TRAPIL possède une large gamme de racleurs XTraSonic, permettant l'inspection des pipelines de différents diamètres, de 8 à 32 pouces. Même si leur conception peut légèrement varier d'un diamètre à l'autre (notamment dans le but de pouvoir passer certains cintres serrés, ou selon l'autonomie requise), leur fonctionnement général reste le même et nous pouvons les considérer similaires dans l'optique de nos travaux.

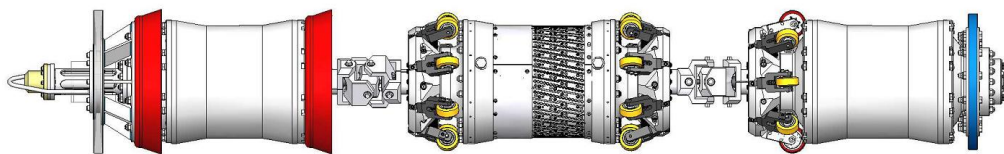


FIGURE 1.1 – Exemple de racleur XTraSonic utilisé par TRAPIL. Ici, un XTraSonic 16 pouces 512 sondes

La Fig. 1.1 présente un exemple de racleur XTraSonic. Il est composé de trois fûts, comprenant de gauche à droite :

- Le fût PC contenant l'ordinateur de bord du racleur contrôlant les sondes ultrasons, ainsi que les disques durs ;
- Le fût sonde où sont disposées les sondes ultrasons sur toute la circonférence ;
- Le fût batterie contenant l'alimentation du racleur.

Le tir des sondes ultrasons est déclenché à l'aide de roues de comptage, symbolisées en rouge sur le fût de droite dans la Fig. 1.1, selon un pas d'acquisition variant selon le diamètre et la longueur du pipeline à analyser. Par exemple les diamètres inférieurs à 16 pouces usent d'un pas d'acquisition de 1.5 millimètres, alors que les diamètres supérieurs à 16 pouces usent d'un pas d'acquisition de 3 millimètres, afin de limiter la quantité de données à stocker. Le nombre de sondes disposées autour du fût sonde dépend également du diamètre du pipeline à inspecter. La correspondance entre diamètre et nombre de sondes est détaillée dans la Table 1.1.

Le racleur est inséré dans le pipeline durant le service et est propulsé par le fluide à l'aide de ses coupelles (en rouge sur le fût de gauche). Les roues normales (en jaune)

diamètre (pouces)	nombre de sondes	pas d'acquisition (mm)
8 à 10	160	1.5
10 à 16	224 ou 256	1.5
16 à 34	512	3

TABLE 1.1 – Table de correspondance entre le diamètre de la ligne, nombre de sonde du racleur et pas d'acquisition

disposées tout autour, liées à un système de ressort, veillent à maintenir le racleur au centre du pipeline afin d'assurer la bonne acquisition des sondes. De plus ces dernières sont inclinées pour que le racleur tourne par pas de trois degrés au fil de son avancée, décrivant une hélice de sorte que, couplé à la disposition en biais des sondes, un recouvrement soit fait pour renforcer l'acquisition, mais également dans le but d'user les coupelles uniformément.

Les sondes ultrasons tirent en L0 (cela veut dire que les sondes émettent une onde longitudinale perpendiculairement à la surface interne du pipeline), et permettent l'acquisition de plusieurs valeurs. La Fig. 1.2 présente les différentes valeurs stockées pour chaque tir de sonde. L'amplitude du signal et le temps de trajet du faisceau ultrason dans le couplant jusqu'à la paroi interne du pipeline sont les premières valeurs à être stockées, nommées respectivement A1 et T1. Ensuite, l'amplitude du signal et le temps de trajet du faisceau ultrason dans le métal entre la paroi interne et la paroi externe sont également stockés et nommés respectivement A2 et T2. Par la suite, les temps de trajet, couplés aux vitesses de propagation du faisceau dans le couplant ou le métal (valeurs invariantes et connues) permettent de déduire la distance entre le racleur et la paroi interne, appelée colonne d'eau, et la distance entre la paroi interne et la paroi externe, donc l'épaisseur du tube. C'est pourquoi par la suite nous nous référerons aux T1 et T2 comme des valeurs en millimètre et non pas en microsecondes.

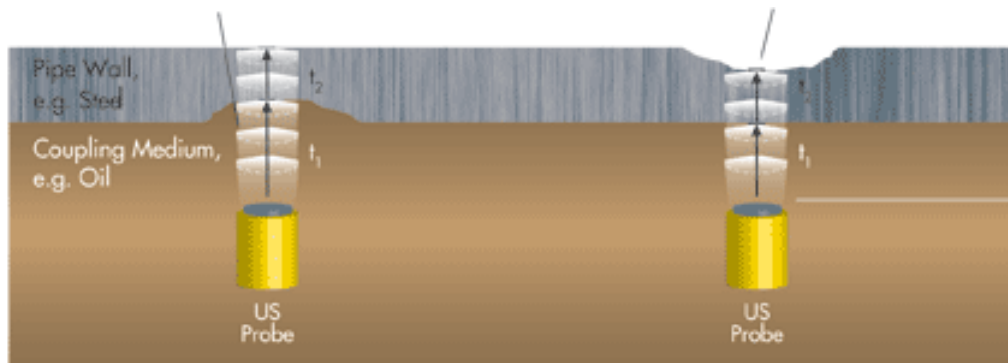


FIGURE 1.2 – Schématisation du tir ultrason pendant le service

La mise en relation de toutes les acquisitions des différentes sondes, permet l'obtention d'une cartographie complète du pipeline, qui peut ensuite être analysée. Les axes de cette cartographie sont, en abscisse les Points Kilométriques (PK) correspondant aux différentes acquisitions selon le pas défini, et en ordonnée la Position Angulaire (PA) correspondant aux différentes sondes positionnées sur la circonférence du racleur. La résolution du PK est directement déduite du pas d'acquisition utilisé par la roue de comptage, tandis que celle du PA correspond à la distance entre deux sondes sur le racleur, environ 5 millimètre. La Fig. 1.3 montre un exemple de cartographie obtenue. On peut y voir une signature de soudure

longitudinale orientée à 3 degrés environ, résultante directe de la rotation du racleur explicitée plus haut. Quatre cartographies sont accessibles à l'analyse, correspondant aux quatre données enregistrées sur chaque tir : A1, A2, T1 et T2 (sur la Fig. 1.3, une cartographie T2). Il est important de noter que l'image obtenue est circonférentielle, le haut de l'image étant voisin du bas.

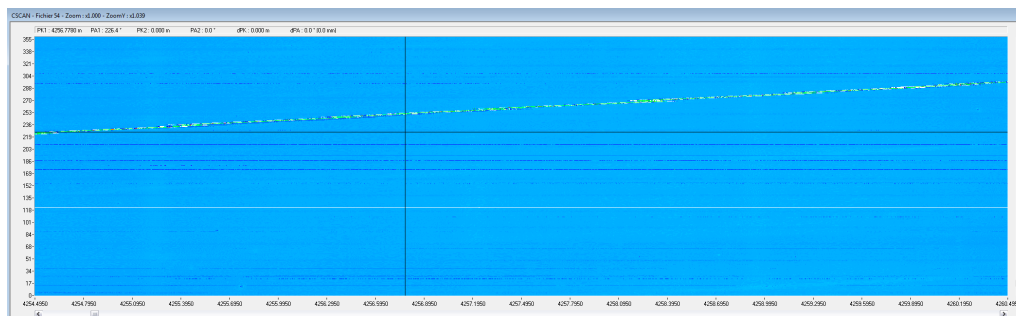


FIGURE 1.3 – Exemple de cartographie T2 sur un tube sain, à soudure longitudinale

Le nombre de données à traiter dépend donc du diamètre de la ligne (donc du nombre de sondes du racleur  $S$ ), de sa longueur  $L$  et du pas d'acquisition  $k$ . Le nombre de pixels  $N$  à traiter d'une ligne est donc de :

$$N = 4 \times S \times \frac{L}{k} \quad (1.1)$$

Par exemple, pour une ligne faisant 15 kilomètres, inspectée par un racleur 256 sondes avec un pas d'acquisition de 1.5 millimètres, le nombre de pixels à traiter est donc d'environ 10 milliards en prenant en compte les quatre sources de données A1, A2, T1 et T2.

## 1.2 Problématique de détection de défauts structurels

Une fois le passage achevé et les données récupérées, l'analyse de la ligne peut débuter. Une phase de formatage est réalisée pour transformer les données brutes en données lisibles par le logiciel d'analyse. Les données sont alors étudiées rapidement afin d'avoir une idée de la qualité globale de l'acquisition, et si les indicateurs de qualité (niveau de bruit, nombre de sondes défectueuses,...) sont acceptables, les analystes commencent par annoter sur le pipeline les différents éléments de ligne afin de pouvoir se repérer, puis parcourent le passage intégralement afin de détecter et d'encadrer les défauts présents.

### 1.2.1 Les soudures d'aboutage et les différents types de tubes

La première phase de l'analyse consiste à détecter et annoter les soudures d'aboutage reliant les différents tubes composant le pipeline. La Fig. 1.4 présente une soudure d'aboutage ainsi que la signature correspondante acquise par le racleur. Découper ainsi le pipeline en tube est important car les tubes composant le pipeline peuvent avoir des caractéristiques différentes (épaisseur, procédé de fabrication). De plus une soudure entraîne une chauffe du métal et l'affaiblit, d'où la définition d'une Zone Affectée Thermiquement (ZAT) de part et d'autre de la soudure où les critères de sélection de défaut doivent être plus drastiques en raison de la fragilisation du métal.

Il existe trois méthodes de fabrication de tubes de pipeline :

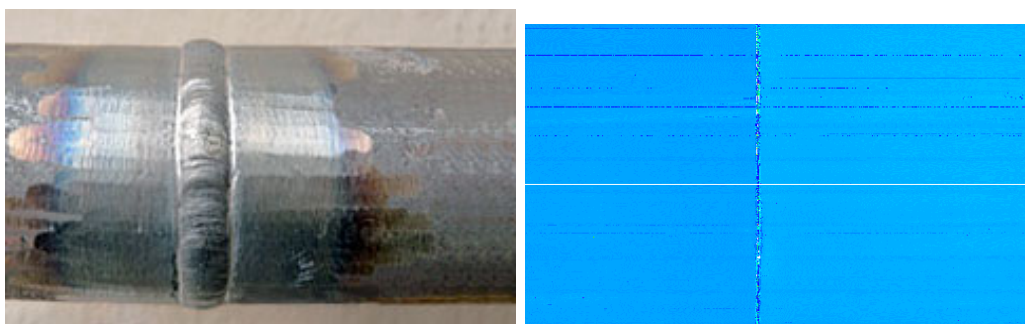


FIGURE 1.4 – Exemple de soudure d'aboutage saine. A gauche une photo terrain, à droite une signature T2 acquise par le racleur

- Les tubes extrudés sont produits par procédé thermomécanique dans lequel une pièce de métal est contraint de traverser une filière ayant la section de la pièce à obtenir. Ces tubes sont réalisés sans soudure, mais le processus de fabrication entraîne des irrégularités de texture sur les surfaces interne et externe.
- Les tubes à soudure longitudinale sont réalisés à partir d'une feuille de métal pliée et soudée dans le sens de la longueur. Une soudure est présente mais la texture est plus nette que les tubes extrudés.
- Les tubes à soudure hélicoïdale sont réalisés de manière similaire aux tubes à soudure longitudinale à l'exception du fait que la soudure est réalisée en hélice et non plus selon la longueur.

### 1.2.2 Les différents éléments de ligne

On appelle élément de ligne tout objet ou partie de la ligne entraînant une signature particulière sur l'inspection ultrasonique sans pour autant être un défaut. On distingue plusieurs catégories :

- Les sections de ligne, rassemblant les tubes aux caractéristiques particulières. On y trouve les cintres, qui sont des coudes dans le pipe, servant à modifier la direction de la ligne. Ils peuvent être d'angle variés, et entraînent des signatures de trainées verticales tout au long du coude. Les cintres sont généralement accompagnés de pertes d'écho car le racleur est bousculé en raison du transfert de masse et l'acquisition est donc difficile. On trouve également dans cette catégorie les gares racleur, servant à insérer ou récupérer le racleur, et qui sont situées bien évidemment en début et en fin d'acquisition. Les joints isolants sont de petites portions servant à isoler électriquement le pipe, généralement lors d'une traversée de fleuve. Enfin, les vannes sont des portions utiles à l'exploitation de la ligne servant à bloquer le flux afin d'arrêter le transport. La Fig. 1.5 présente les signatures de ces différents éléments acquises par le racleur.
- Les déviations, qui sont des croisements, généralement un tube se séparant en deux (ou deux tubes se rejoignant). On trouve dans cette catégorie les tés, déviations simples en forme de T, les *split tee* qui sont en général des tés fabriqués en ajoutant une coquille autour du tube avant de le percer directement, et enfin les piquages qui sont des tés de petits diamètres, généralement utilisés pour faire des prélèvements, des purges ou des dégazage. La Fig. 1.6 présente les signatures acquises par le racleur

de ces différents éléments.

- Les éléments de réparation, enfin, sont comme leur nom l'indiquent des éléments qui ont été rajoutés sur le pipeline afin de réparer un défaut préalablement détecté. On trouve dans ce groupe les coquilles, deux demis tubes de renforts apposés par dessus un tube fragilisé et soudés à ce dernier ; les manchons, identiques aux coquilles mais qui viennent réaliser ou renforcer la jonction entre deux tubes ; les patches enfin sont une pièce de métal soudé directement par dessus un défaut afin de le renforcer. La Fig. 1.7 présente les différentes signatures obtenus par acquisition racleur de ces éléments.

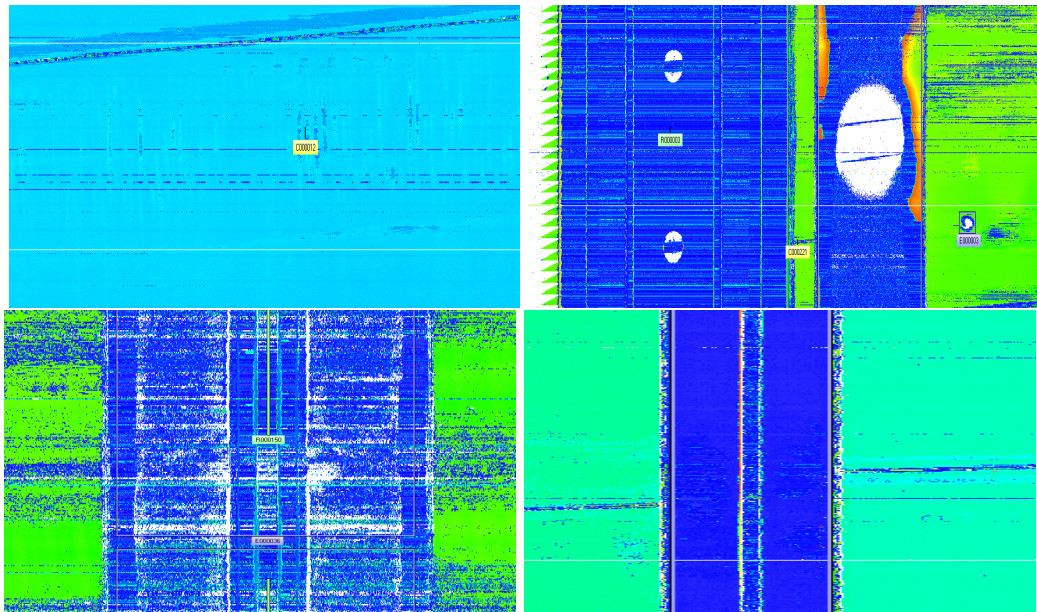


FIGURE 1.5 – Exemples de signatures de section de ligne : en haut à gauche un cintre, en haut à droite une gare racleur, en bas à gauche une vanne, et en bas à droite un joint isolant.

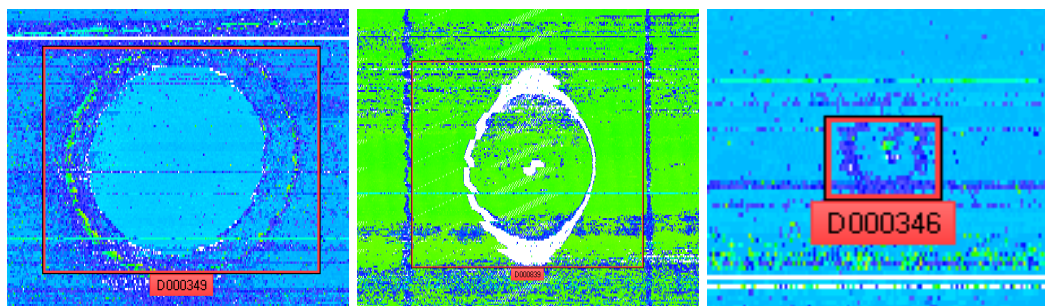


FIGURE 1.6 – Exemples de signatures de déviation : de gauche à droite, un té, un split tee, et un piquage.



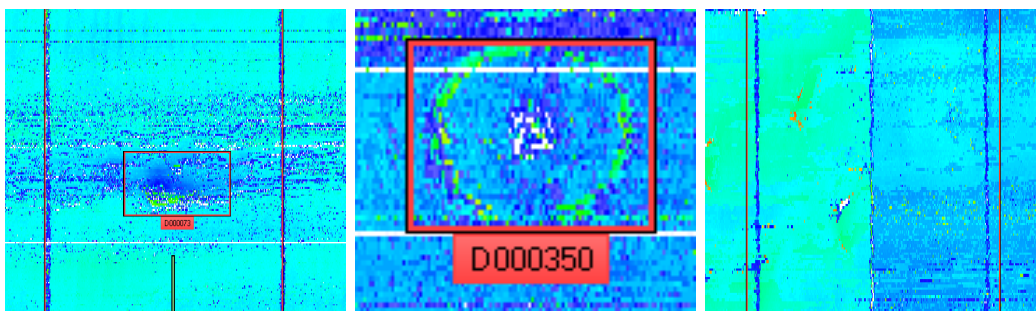


FIGURE 1.7 – Exemples de signatures de réparation : de gauche à droite, une coquille, un patch, et un manchon.

### 1.2.3 Les différents types de défauts

Le Pipeline Operators Forum (POF) [7] a classifié les différents défauts existants dans diverses catégories. La table 1.2 les présente ainsi que la définition qui leur est communément associée.

Il est à noter que l'inspection ultrasonique L0 ne permet pas de détecter les fissures car celles-ci sont perpendiculaires à la section du tube. Ces types de défauts ne rentrent donc pas en compte dans notre problématique. Quatre familles de défauts sont prédominantes et sont généralement utilisées lors des inspections :

- Les délaminages ;
- Les corrosions ;
- Les défauts géométriques (enfoncement et ovalisation) ;
- Les manques de métal (Coup d'arc, griffure, meulage, écaillage).

La Fig. 1.8 présente un exemple de chaque type de défaut et les signatures correspondantes obtenues par le racleur.

Il s'agit dans la très grande majorité des cas de défauts naturels, et par conséquent leurs caractéristiques sont extrêmement variables (taille, profondeur, forme, texture résiduelle). Il est donc très difficile de les caractériser de manière générale et bien souvent la frontière visuelle entre deux familles de défauts s'en trouve très difficile à définir, surtout lorsque s'ajoute le bruit d'acquisition par dessus.

### 1.2.4 Les erreurs d'acquisition et l'inhibition

L'acquisition ultrasonique est une technologie sensible et malgré les efforts de l'entreprise pour stabiliser le racleur dans un passage, des erreurs d'acquisition peuvent apparaître. Malfunction des sondes, glissement de la roue PK, mouvement brusque du racleur sont autant de raisons pouvant entraîner l'apparition de données erronées. L'acquisition peut alors être très difficile à analyser car ce bruit recouvre les informations réelles.

De plus les limitations techniques font que le racleur est incapable de mesurer les épaisseurs inférieures à 3 millimètres. Si le cas se produit, la valeur stockée est placée à 3.1 millimètres mais il s'agit d'une valeur incertaine ; on parle dans ce cas de valeur d'inhibition. Ces valeurs arrivent régulièrement lorsque le racleur est bousculé ou que la surface du tube est en peau d'orange, l'acquisition est alors extrêmement perturbée, comme on peut le voir dans la Fig. 1.9.

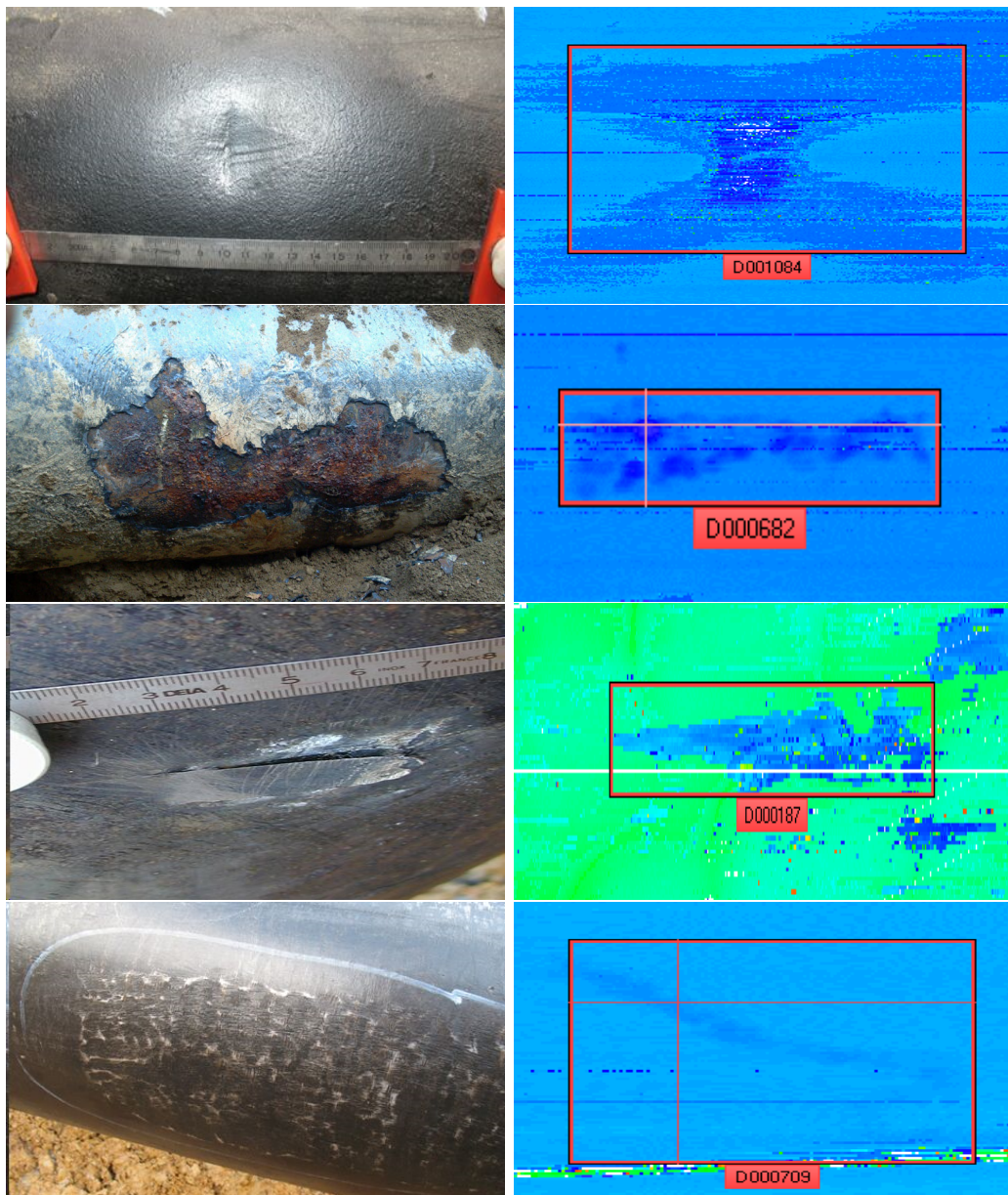


FIGURE 1.8 – Exemples d’acquisition pour chacune des quatre familles majeures de défauts. A gauche une photo terrain, à droite, une signature obtenue par le racleur sur ce type de défaut. De haut en bas : Enfoncement, Corrosion, Délaminage, Sous-épaisseur.

### 1.2.5 Le processus d’analyse manuelle

L’équipe d’analystes récupère les données extraites du racleur et les visualisent à l’aide de leur logiciel de visualisation (LARA). Une fois les soudures d’aboutage annotées, ils parcourent la ligne tube par tube et relèvent les défauts qu’ils observent avant de les encadrer et de les identifier selon leur type. Par la suite chaque défaut est étudié afin de déterminer la criticité de celui-ci. La somme de toutes ces informations est ensuite compilée dans un rapport rendu au client, laissant la charge à celui-ci de réaliser les réparations qui lui semblent nécessaires. Le délai demandé entre la réception des données et la remise

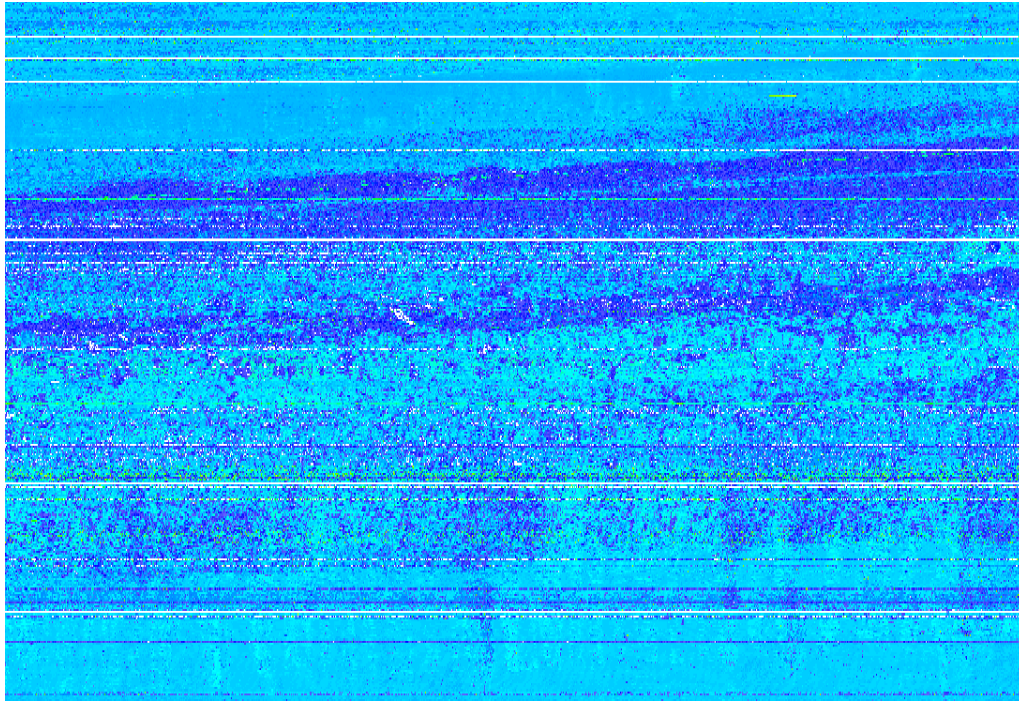


FIGURE 1.9 – Exemples de tube extrêmement bruité, auquel s’ajoute plusieurs sondes défectueuses (lignes blanches).

du rapport est d’environ une semaine par personne pour dix kilomètres, plus une semaine pour réaliser le rapport. Il faut donc quatre semaine/homme pour analyser une ligne de trente kilomètres. Dans le but de réduire ce temps, TRAPIL a donc souhaité automatiser une partie de son processus d’analyse afin de gagner en réactivité.

### 1.3 Objectifs du projet

L’objectif principal du projet est assez vaste car il s’agit d’apporter, à l’aide d’outils de traitement d’image et de signal, une aide à l’analyse manuelle dans le but de faire gagner du temps et de la réactivité au processus de détection et d’identification des défauts. L’automatisation de ce dernier rencontre de nombreuses difficultés :

- La quantité de données à traiter est gigantesque et la complexité des méthodes doit être maîtrisée, barrant la route à la plupart des outils de traitement d’image complexes (Champs de Markov, Contours actifs,...).
- Les items à détecter sont d’origine naturelle, donc possèdent des caractéristiques extrêmement variées en taille, en forme et en texture. La création d’un modèle est donc difficile.
- L’acquisition par le racleur est imparfaite, dégradant la qualité des données à analyser (bruit, sondes defectueuses...).
- L’état du pipeline n’est pas toujours optimal, des irrégularités de fabrication peuvent entraîner des signatures anormales sans pour autant être des défauts, et des résidus de matière peuvent se déposer au fond du tube, affectant l’acquisition.

L'analyse automatique n'a donc pas pour but de se substituer à l'analyse manuelle fine, mais de venir en renforcement de celle-ci. L'objectif est donc de réaliser une pré-analyse permettant un réel gain de temps, relevant le plus grand nombre de défauts possible sans retourner un trop grand nombre de fausses alarmes. Cette première analyse pourra être vérifiée rapidement par un analyste, qui se contentera de passer en revue uniquement les retours de la méthode, et ainsi fournir au client un premier rapport incomplet mais contenant les défauts les plus notables, qu'il pourra commencer à investiguer pendant que l'analyse standard sera réalisée.

Les différentes étapes de l'analyse automatique suivra donc le même processus que les experts : En premier lieu la ligne sera segmentée en tube via une détection de soudures d'aboutage, puis les tubes seront analysés un par un afin de détecter puis d'identifier les signatures anormales. En raison de leur criticité et de la difficulté d'analyse de ces zones, couplées au fait qu'elles représentent une part faible de la quantité de données totale, les ZAT ont été considérées comme n'entrant pas dans notre problématique et ne sont donc pas étudiées.

## 1.4 Etat de l'art sur la problématique

### 1.4.1 Méthodes utilisées pour ce domaine d'application

Depuis longtemps les méthodes de traitement d'image et du signal sont utilisées dans l'industrie du médical ou de la sécurité. Cependant, dans le domaine du contrôle non destructif ces disciplines sont encore jeunes et les entreprises commencent seulement à s'y intéresser. La grande majorité de la recherche concerne encore les méthodes d'acquisition et bien peu sont orientées sur le traitement en lui-même. Nous savons que plusieurs entreprises internationales réalisant des inspections racleur ultrasonique pour les pipelines d'hydrocarbure utilisent de telles méthodes d'automatisation, mais leur détail n'a pas été publié et nous n'avons pas de vision sur leurs performances, car elles sont toujours couplées à l'analyse manuelle. Par ailleurs, nous n'avons d'information que sur le résultat final incorporant l'intégralité du processus. Il existe en revanche des domaines d'applications assez proches ayant fait l'objet de publications, et nous allons nous orienter par défaut sur ces dernières.

### 1.4.2 Méthodes utilisées pour des problématiques similaires

Il est possible de trouver des publications traitant du contrôle non destructif appliqué aux pipelines mais la grande majorité de la recherche est orientée sur les racleurs MFL (*Magnetic Flux Leakage*), car 90% des inspections dans le monde sont réalisées grâce à cette technologie qui possède l'avantage de fonctionner aussi bien dans le liquide que dans le gaz, ouvrant donc des possibilités de marché aux entreprises spécialisées, là où la technologie ultrason ne peut fonctionner que dans le liquide. Des techniques d'apprentissage très variées ont été développées dans le cadre de cette problématique, comme la transformée en ondelette discrète [21] ou encore les réseaux de neurone [5]. Cependant les données acquises sont de nature très différentes et même si l'application est similaire la problématique elle, est incomparable.

De nombreuses publications proposent d'autres méthodes d'acquisition et de traitement pour le contrôle non destructif des pipelines ([28], [26], [15], [32]) mais leur technologie est encore une fois bien différente de l'ultrason et les images résultantes sont donc trop

différentes. La plupart des publications traitant véritablement de racleurs ultrasons sont portées sur les pipelines de béton [11] car dans ce cas le MFL ne s'applique pas. Cependant la nature des défauts recherchés dans cette application est également différente car dans ce type de pipeline seules les grosses fissures sont dangereuses. Enfin, de nombreuses publications sont orientées sur la technologie ultrasonique multi-éléments (*Phased Array*) comme dans [13], ou étudient le comportement de ces faisceaux de sondes ([10] ou [25]). Mais cette technologie est encore au stade expérimental et n'est donc pas encore ou peu utilisée dans l'industrie.

Néanmoins à ce jour, la méthode de détection/reconnaissance de défauts faisant référence reste l'approche manuelle. Dans les chapitres suivants, nous proposons une chaîne de traitement complète comprenant la détection de zones à risque, leur segmentation et leur identification selon les classes de défauts utilisées à ce jour par les analystes. Le chapitre suivant introduit en particulier une méthode automatique de détection des soudures, première étape nécessaire au traitement des données dites "en coeur de tube".

Dénomination	Définition
Coup d'arc	Point où le métal a été mis en fusion superficiellement par un arc électrique
Défaut Artificiel	Défaut usiné artificiellement dans un but de test
Flambage	Rupture partielle du tube due à une torsion excessive, causée par une modification brutale de l'environnement (ex : Tremblement de terre)
Corrosion	Réaction électrochimique entre le pipeline et l'environnement, entraînant une perte de métal
Zone de corrosion	Corrosions multiples sur une large zone
Fissure	Fracture plane (2 dimensions) avec possible déplacement de métal, dont la cause est inconnue
Enfoncement	Distorsion de la surface du pipeline entraînant une modification locale du diamètre interne
Enfoncement avec manque de métal	Enfoncement ayant entraîné une perte de métal locale
Griffure	Perte de métal induite mécaniquement entraînant la formation de tranchées
Meulage	Réduction d'épaisseur du pipeline réalisée à la main ou par disquetteuse
Fissure sur soudure d'aboutage	Fissure localisée sur ou à proximité d'une soudure d'aboutage
Défaut sur soudure d'aboutage	Défaut localisé sur ou à proximité d'une soudure d'aboutage
Fissuration assistée par hydrogène	Fissure causée par atomes d'hydrogène
Délaminage	Imperfection ou discontinuité du métal causant une séparation en couche parallèle ou angulaire au tube
Fissure sur soudure longitudinale	Fissure localisée sur ou à proximité d'une soudure de fabrication longitudinale
Défaut sur soudure longitudinale	Défaut localisé sur ou à proximité d'une soudure de fabrication longitudinale
Ovalisation	Déformation généralisée du tube
Défaut de fabrication	Anomalie créée lors de la fabrication du tube (ex : particule étrangère dans le métal)
Zone de défaut de fabrication	Anomalie de fabrication repertoriée sur une large zone
Fissuration par corrosion sous contrainte	Fissure créée par la mise sous contrainte d'une zone corrodée
Ecaillage	Abrasion de la surface du pipeline entraînant une texturation du métal
Fissure sur soudure hélicoïdale	Fissure localisée sur ou à proximité d'une soudure de fabrication hélicoïdale
Défaut sur soudure hélicoïdale	Défaut localisé sur ou à proximité d'une soudure de fabrication hélicoïdale
Repliure	Déformation extrême du métal entraînant la formation d'un pli
Autre	Tout défaut détecté n'appartenant pas à l'une des catégories ci-dessus

TABLE 1.2 – Liste des types de défauts caractérisés par le POF



## Chapitre 2

# Détection des soudures d'aboutage

### 2.1 Introduction

La détection des soudures d'aboutage est la première étape du processus d'analyse. Cette dernière est indispensable afin de permettre la segmentation de la ligne en ses différents tubes, permettant un repérage plus aisé mais aussi afin de faciliter les étapes suivantes de la chaîne de traitement. Si une ligne mesure plusieurs kilomètres de long, un tube lui ne dépasse jamais seize mètres, simplifiant la mise en mémoire des données et permettant une analyse plus fine de chacun d'entre eux car la topographie varie d'un tube à l'autre mais jamais dans un même tube.

Les difficultés inhérentes à la détection de soudure d'aboutage sont de cinq natures :

- La taille des images à traiter (plusieurs Giga pixels) ;
- Les images peuvent être déformées par un bruit intense et/ou des pertes de données ;
- De par la nécessité d'automatisation, il s'agit d'un problème à traiter en aveugle : il n'existe pas de signatures références à rechercher ;
- Les soudures possèdent des caractéristiques variables en largeur et en texture ;
- Les procédés de fabrication de tubes peuvent entraîner des corruptions de texture, même en l'absence d'anomalies.

Les méthodes classiques de traitement d'image comme les détection de contours [12], les champs de Markov [33] ou bien encore les modèles par contour actif [16] ne peuvent pas être utilisés pour détecter les soudures d'aboutage dans ce contexte principalement à cause de leur complexité algorithmique inadéquate à la quantité de données à traiter. De plus ces méthodes nécessiteraient de profondes adaptations et des pré-traitements pour pouvoir gérer le bruit structurel et les données manquantes en raison de la possible défaillance des sondes.

La méthode que nous proposons, nommée *READ-weld*, répond aux cinq problèmes identifiés ci-dessus et est détaillée par la suite. Synthétiquement, dans un premier temps, nous identifions la problématique de détection d'événements ciconférentiels et établissons les notations clés ; puis nous présentons le détail de la méthode avant de donner les résultats obtenus au travers de notre processus expérimental, en particulier en la comparant à la méthode automatique, développée par une société externe, que TRAPIL utilisait jusqu'à présent.

Dans cette section, et afin d'illustrer nos propos, nous considérons une portion particulière d'une ligne inspectée par un racleur 224 sondes, avec un pas de 1.5 millimètres, disposant d'une soudure d'aboutage en son centre, fortement bruitée sur sa partie gauche



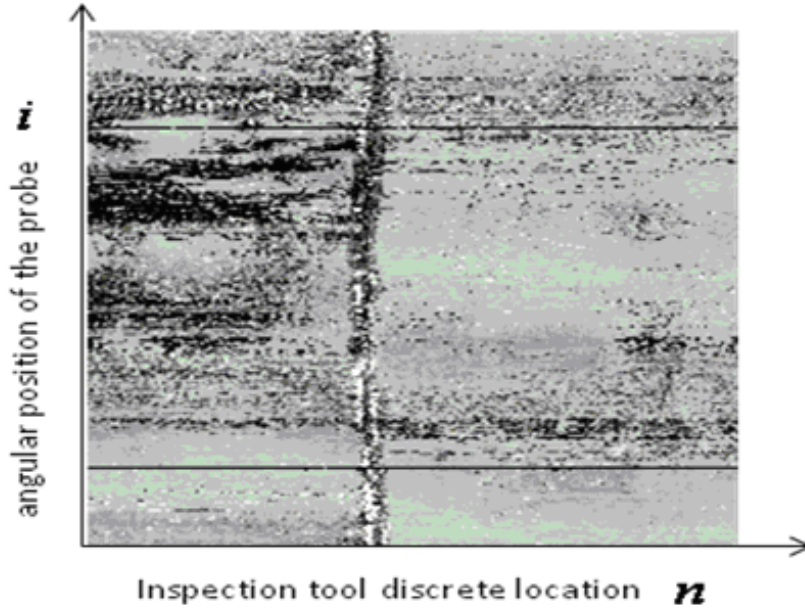


FIGURE 2.1 – *Portion de ligne utilisée comme exemple dans cette partie. On y voit une signature de soudure au centre et un bruit intense à sa gauche rendant sa détection difficile*

en particulier, rendant sa détection difficile même à l'oeil nu. Cette section type est illustrée au moyen de la Figure 2.1.

## 2.2 Problématique de détection d'événements circonférentiels

Nous appelons FCE (*Full Circumferential Event*) une signature particulière et répercutée sur l'ensemble de la circonférence du tube (ce qui est le cas pour les soudures d'aboutage). A l'inverse, nous appelons PCE (*Partial Circumferential Event*) une signature anormale qui ne serait pas répercutée sur l'ensemble de la circonférence (un défaut ponctuel, une trainée de bruit, les signatures de rainures entraînées par les cintres). Le problème de détection des FCE est un problème à trois hypothèses [30] tel que,

$$\begin{aligned} H_0 &: y_i[n] = w_i[n] \\ H_z &: y_i[n] = x_i^z[n] + w_i[n], z \in \{FCE, PCE\} \end{aligned} \quad (2.1)$$

Dans l'équation (2.1),  $w_i[n]$  désigne le bruit d'acquisition acquis par la sonde  $n$  à l'indice (PK)  $i$ , alors que  $x_i^z[n]$  dénote le signal géométrique mesuré par la sonde  $i$  provenant d'un FCE ( $z = FCE$ ) ou d'un PCE ( $z = PCE$ ) localisé à l'abscisse  $n$ . Il est important de rappeler qu'en raison de sa nature automatique le problème auquel nous faisons face est aveugle :  $x_i^z[n]$  n'est pas connu et ne dispose d'aucun modèle mathématique établi dans la communauté des inspections de pipelines. Nous définissons l'opérateur de contraste  $C_{a,b}^\pi(\cdot)$  entre deux hypothèses  $H_a$  and  $H_b$  tel que,

$$C_{a,b}^\pi(\cdot) = \frac{|E[\cdot|H_a] - E[\cdot|H_b]|^2}{\pi \cdot Var[\cdot|H_a] + (1 - \pi) \cdot Var[\cdot|H_b]} \quad (2.2)$$

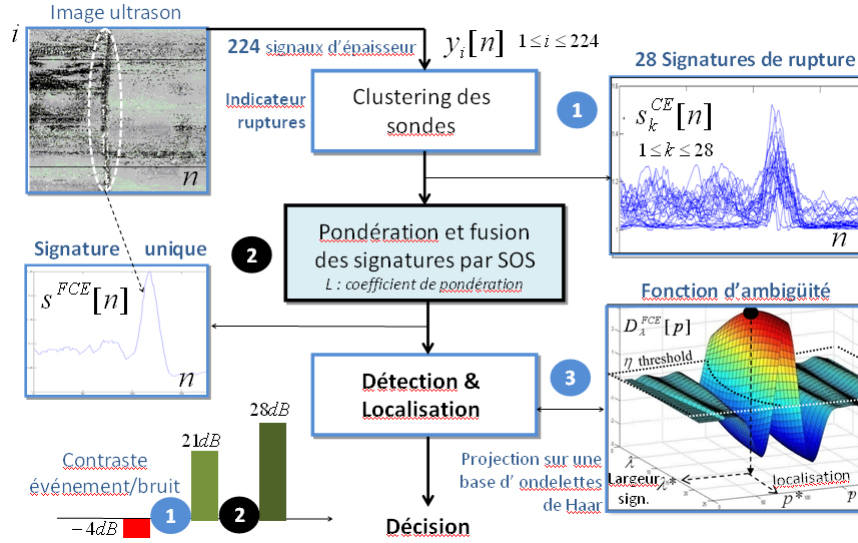


FIGURE 2.2 – Schéma méthodologique de READ-weld

Dans l'équation (2.2), nous avons  $(a, b) \in \{0, FCE, PCE\}^2$  et  $0 \leq \pi \leq 1$ . Le contraste d'origine entre le bruit et un FCE,  $C_{FCE,0}^0(y_i[n])$  peut facilement atteindre une valeur typique de  $-4\text{dB}$ . Cela signifie que les pixels appartenant à  $H_{FCE}$  et le bruit  $H_0$  ne peuvent être distingués sans un procédé d'amélioration des signatures de FCE et une destruction du bruit environnant.

## 2.3 Méthode proposée : *READ-weld*

### 2.3.1 Vue d'ensemble de *READ-weld*

*READ-weld* procède en trois étapes distinctes comme synthétisé dans la Figure 2.2.

- La première étape élabore quelques fonctions de signatures distribuées  $s_k^{CE}[n]$ ,  $1 \leq k \leq N_c$ , ( $N_c = 28$ ) chacune contenant des caractéristiques de texture d'événements circonférentiels. Chaque signature est obtenue après collecte des mesures obtenues sur un sous-ensemble (appelé *cluster*) des sondes disponibles (typiquement 4 sondes par *cluster*). Par conséquent seul un nombre compressé et définissable de sondes restent impliquées (ici 112 au lieu de  $N_s = 224$  au départ).
- La seconde étape combine et pondère les signatures CE en une unique signature FCE,  $s^{FCE}[n]$  améliorée ou "circonférentiellement orientée".
- La troisième étape enfin élabore une fonction d'ambiguïté [24] afin de détecter et de localiser les FCE.

### 2.3.2 Première étape de *READ-weld* : Compression et signatures CE

- La première étape comprend elle même trois sous-étapes. Premièrement l'image brute subit un filtrage médian circonférentiel (donc dans notre problématique, vertical) afin d'effectuer un premier débruitage, tout en préservant l'information de changement horizontal. On denote  $\tilde{y}_i[n]$  les données ainsi débruitées.

- Au cours de la seconde étape,  $N_c$  clusters  $Q_k = \{ i_1^k \leq i \leq i_q^k \}$  de  $q$  sondes sont construits. Pour des considérations de compression, les  $N_c$  clusters sont peuplés avec un nombre réduit de sondes,  $qN_c \leq N_s$  (les sondes sont sélectionnées à la manière d'un sous-échantillonnage, donc si  $q = 0.5$ , alors une sonde sur deux sera prise en compte). Le choix du nombre de clusters ainsi que le nombre de sondes par cluster a été étudié *a posteriori* en comparant les résultats de gain de contraste pour tous les cas sur 300 soudures, et il en résulte que des résultats optimaux sont obtenus pour  $N_c = 7q$ . Ainsi dans notre exemple avec un racleur 228 sondes, réduit à 114 par l'étape de compression ( $q = 0.5$ ), le choix optimal est de prendre  $N_c = 28$  et  $q = 4$ .
- Troisièmement, dans le but d'améliorer le contraste entre le bruit et les événements CE, tout en réduisant encore plus la dimensionnalité des données de  $qN_c$  à  $N_c$ , les clusters de signaux, i.e.  $\tilde{y}_i[n], i \in Q_k$ , subissent un traitement nommé "détection de rupture" [1] prenant en compte les signatures CE  $s_k^{CE}[n], 1 \leq k \leq N_c$  avec,

$$s_k^{CE}[n] = \frac{\mu_{k,short}[n]}{\mu_{k,long}[n]} \quad (2.3)$$

Dans l'équation (2.3),  $\mu_{k,long}[n]$  et  $\mu_{k,short}[n]$  désignent respectivement les moyennes mouvantes des données débruitées à long terme et à court terme, au sein du cluster  $Q_k$ ,

$$\mu_{k,long}[n] = \frac{1}{qT_{long}} \cdot \sum_{i \in Q_k} \sum_{l=1}^{T_{long}} \tilde{y}_i[n-l] \quad (2.4)$$

$$\mu_{k,short}[n] = \frac{1}{qT_{short}} \cdot \sum_{i \in Q_k} \sum_{l=0}^{T_{short}-1} \tilde{y}_i[n+l] \quad (2.5)$$

Le contraste moyen,  $\langle C_{FCE,0}^0(s_k^{CE}[n]) \rangle_{1 \leq k \leq N_c}$  entre les FCE et le bruit, des signatures CE décrites par l'équation (2.3) est égal à 21dB, atteignant ainsi un gain de 25dB vis-à-vis des images brutes. Par conséquent, comme montré dans la Figure 2.3, une marque floue, mais partagée par la plupart des signatures CE (*en présence de FCE*) prend forme.

### 2.3.3 Seconde étape de *READ-weld* : Signature unique FCE améliorée

L'objectif de la seconde partie est de rassembler les signatures CE  $s_k^{CE}[n], 1 \leq k \leq N_c$  en une unique signature FCE  $s^{FCE}[n]$  (*atteignant ainsi une compression maximale*) qui sera "orientée circonférentiellement". Cette dernière caractéristique implique que la combinaison des signatures CE  $s_k^{CE}[n]$  pour créer  $s^{FCE}[n]$  devra renforcer une alerte particulière *seulement* si elle est présente sur la *totalité* des signatures, mettant en avant une alerte totalement circonférentielle, FCE. A l'inverse, si seules quelques signatures CE présentent une alerte (*comme on peut s'y attendre en cas de PCE*), la méthode de combinaison devra "aplatir" le signal présent dans les autres signatures. Formellement, une signature FCE  $s^{FCE}[n]$  devra présenter un fort contraste entre un PCE et un FCE et cela afin de réduire la probabilité de fausse détection. En gardant à l'esprit ces quelques contraintes, nous avons,

$$s^{FCE}[n] = \sum_{k=1}^{N_c} \alpha_k[n] \cdot s_k^{CE}[n] \quad (2.6)$$

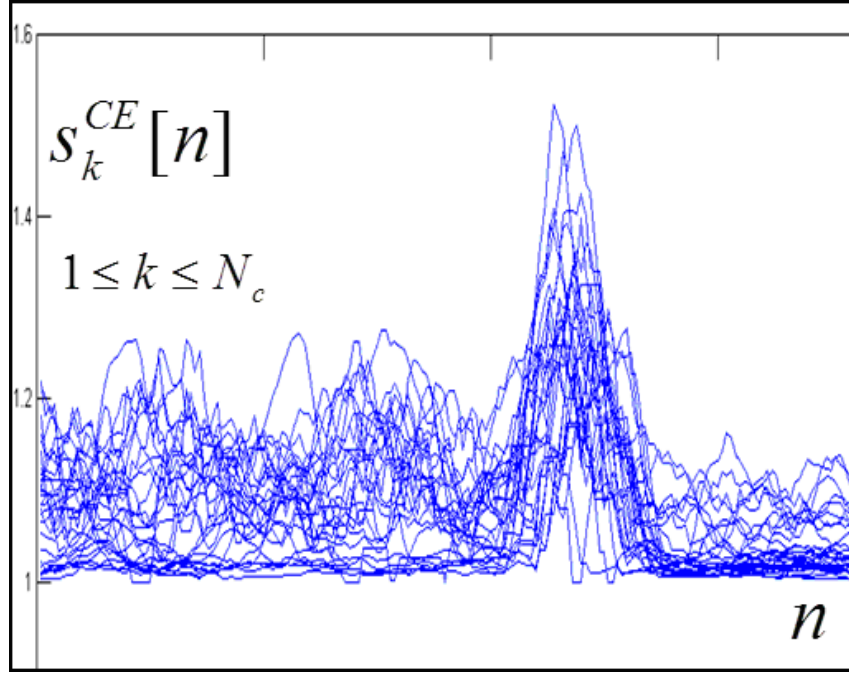


FIGURE 2.3 – Signatures CE superposées  $s_k^{CE}[n]$ ,  $1 \leq k \leq N_c$  par rapport à  $n$ , avec  $N_c = 28$ , après traitement de notre exemple Figure 2.1.

Les fonctions de pondération  $\alpha_k[n]$  sont solution du problème d'optimisation suivant,

$$\alpha_k[n] = \frac{\text{Arg} \left\{ \min_{\alpha_k} \left( \sum_{k=1}^{N_c} (A - \alpha_k[n] \cdot \mu_k^L[n])^2 \right) \right\}}{\sum_{k=1}^{N_c} \alpha_k[n] = 1} \quad (2.7)$$

Dans l'équation (2.7),  $A$  est une constante et  $\mu_k^L[n]$  correspond au moment d'ordre  $L$  de la signature CE  $s_k^{CE}[n]$ , i.e.

$$\mu_k^L[n] = \frac{1}{2M+1} \cdot \sum_{l=-M}^{l=+M} (s_k^{CE}[n+l])^L \quad (2.8)$$

Les fonctions de pondération  $\alpha_k[n]$  visent à égaliser partiellement les moments d'ordre  $L$  des signatures CE. Un calcul trivial nous mène à,

$$\alpha_k[n] = \frac{1}{\mu_k^L[n]} \cdot \frac{1}{\sum_{r=1}^N \frac{1}{\mu_r^L[n]}} \quad (2.9)$$

Il est clair en regardant l'équation (2.9) que si la signature CE est un FCE, tous les moments d'ordre  $L$ ,  $\mu_k^L[n]$ , partageront quasiment la même valeur et les  $\alpha_k[n]$  seront presque identiques, i.e.  $\alpha_k[n] \approx \frac{1}{N_c}$ . Le combineur sera donc *uniforme*. A l'inverse, prenons en compte un PCE qui se répartirait sur la moitié des signatures CE, pendant que l'autre

moitié ne subit qu'un léger signal de fond. Pour mieux comprendre le comportement des fonctions de pondération dans ce dernier cas de figure et pour simplifier, nous assumons par exemple que les moments d'ordre  $L$  du bruit de fond sont tous égaux à une valeur infime dénotée  $\epsilon$ , avec  $\epsilon \ll 1$ . Toujours pour simplifier, nous supposons que l'autre moitié des moments d'ordre  $L$  inhérents au PCE sont tous égaux à une grande valeur notée,  $\frac{1}{\epsilon}$  choisi comme l'inverse de  $\epsilon$ . Du fait de l'hypothèse de l'équation (2.9) les poids du fond notés  $\alpha_\epsilon$ , sont identiques. Même chose pour les poids du PCE notés  $\alpha_{\frac{1}{\epsilon}}$ . La relation suivante, provenant de l'équation (2.9) entre les deux poids s'applique alors :  $\alpha_{\frac{1}{\epsilon}} = \epsilon^2 \cdot \alpha_\epsilon$ . Cette dernière relation implique donc que dans le cas d'un PCE, les signatures élaborées par les clusters de PCE seront drastiquement atténuées au cours du processus de combinaison, car mis en relations avec les signaux de fond. Comme spécifié au début de cette section, une géométrie de PCE entrainera de ce fait une signature FCE où toutes les composantes d'alerte auront pratiquement disparues, réduisant le risque de fausse détection.

La figure 2.4 (a) présente une signature FCE  $S^{FCE}[n]$  par rapport à  $n$ , obtenue après combinaison des 28 signatures distribuées  $S_k^{CE}[n]$  montrées dans la Figure 2.3, en accord avec la combinaison orientée circonférentiellement décrit dans les équations (2.6), et (2.9) avec  $L = 10$ . La figure 2.4 (b) confirme que, dans les équations (2.6) et (2.9) les moments d'ordre supérieur  $L = 10$ , offrent le meilleur contraste  $C_{FCE,PCE}^0(s^{FCE})$  entre les hypothèses FCE et PCE, après combinaison. Au travers de la seconde étape de *READ-weld* le contraste entre les FCE et le bruit continue d'augmenter, atteignant une valeur de  $C_{FCE,0}^0 = 28dB$ . La seconde étape permet donc un gain de  $7dB$  vis-à-vis de l'étape une et un gain de  $32dB$  vis-à-vis de l'image d'origine. A noter que le cas particulier  $L = 0$  correspond à un simple moyennage arithmétique des signatures CE.

### 2.3.4 Troisième étape de *READ-weld* : Détection et localisation

La troisième et dernière étape de *READ-weld* se découpe de nouveau en trois sous-étapes.

- Premièrement, la signature FCE  $s^{FCE}[n]$  subit un filtrage passe-bas ou dérivatif affiné,

$$\tilde{s}^{FCE}[n] = \sum_{l=0}^{T-1} s^{FCE}(n+l) - \sum_{l=1}^T s^{FCE}(n-l) \quad (2.10)$$

$\tilde{s}^{FCE}[n]$  est appelée signature FCE reformée. La dérivation affinée apporte deux choses. Elle aide à éliminer les doubles bosses pouvant parfois apparaître sur certaines signatures FCE provenant d'une soudure particulièrement large (*dans la direction de  $n$ , voir la figure 2.1 par exemple*), avec une géométrie irrégulière. Les doubles bosses pourraient amener à faussement détecter deux FCE. La dérivation contribue également à mettre en forme la signature à détecter (*en mode aveugle*) vis-à-vis des ondelettes de Haar utilisées à la sous-étape suivante.

- Ensuite, une fonction d'ambiguïté [24]  $D_\lambda^{FCE}[p]$  est élaborée en projetant la signature mise en forme  $\tilde{s}^{FCE}[n]$  sur une base d'ondelettes de Haar  $H_\lambda[n]$  [17],

$$D_\lambda^{FCE}[p] = \sum_n \tilde{s}^{FCE}[n] \cdot H_\lambda[n-p] \quad (2.11)$$

La fourchette de valeurs d'échelle,  $\lambda$  est choisie afin de balayer les différentes largeurs possibles de soudures  $\tilde{s}^{FCE}[n]$ .

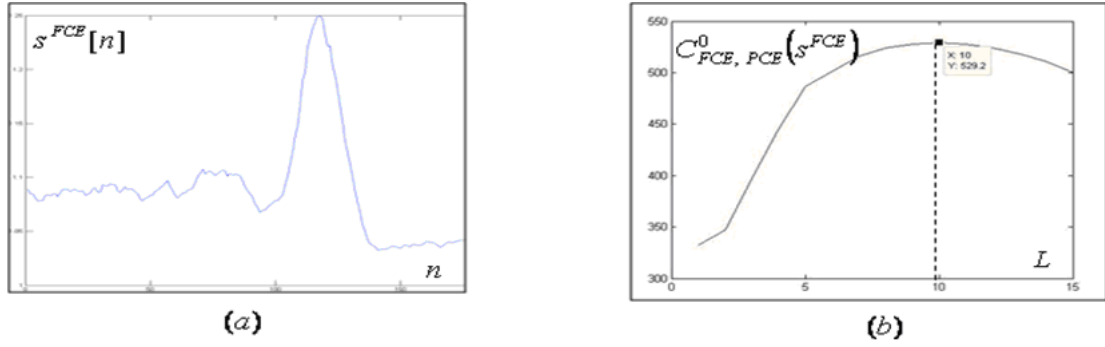


FIGURE 2.4 – (a) Signature FCE  $s^{FCE}[n]$  par rapport à  $n$ , obtenue après combinaison des signatures CE données dans la figure 2.3 au travers de la seconde étape de *READ-weld*. (b) Contraste  $C_{FCE, PCE}^0(s^{FCE})$  de la signature FCE en fonction de l'ordre  $L$  des moments  $\mu_k^L$ , voir Eq. (2.6) et (2.9).

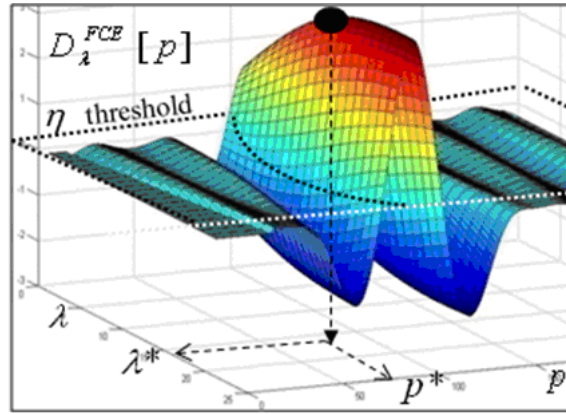


FIGURE 2.5 – Fonction d'ambiguïté  $D_{\lambda}^{FCE}[p]$  basée sur la signature FCE  $s^{FCE}[n]$  donnée dans la figure 2.4 (a), après passage dans les sous-étapes 3.1) and 3.2).

La figure 2.5 présente les variations de la fonction d'ambiguïté  $D_{\lambda}^{FCE}[p]$  par rapport à la valeur d'échelle  $\lambda$  et l'index de localisation  $p$ .

- La valeur d'échelle optimale  $\lambda^*$  est choisie en suivant la règle :

$$\lambda^* = \text{Arg}\{\text{Max}_{\lambda}\{D_{\lambda}^{FCE}[p]\}\} \quad (2.12)$$

Enfin, l'hypothèse  $H_{FCE}$  est décidée si la *fonction de détection*  $D_{\lambda}^{FCE}[p]$  franchit un seuil  $\eta$ . L'index de localisation  $p^*$  du FCE détecté est finalement obtenu en trouvant la localisation du maximum de la fonction de détection, i.e.

$$p^* = \text{Arg}\{\text{Max}_p\{D_{\lambda}^{FCE}[p]\}\} \quad (2.13)$$

La figure 2.5 illustre cette étape.

## 2.4 Performances de la méthode *READ-weld*

### 2.4.1 Probabilités de bonnes détections de FCE et probabilités d'erreurs PCE

Nous définissons la probabilité de bonne détection  $\beta_{FCE}$  de  $H_{FCE}$  comme,

$$\beta_{FCE} = 1 - Pr[H_{\overline{FCE}}|H_{FCE}] \quad (2.14)$$

où  $Pr[H_{\overline{FCE}}|H_{FCE}]$  est la probabilité de manquer  $H_{FCE}$ . Afin de jauger les performances, nous introduisons également la probabilité d'erreur  $\xi_{PCE}$  de décider faussement  $H_{FCE}$  au lieu de  $H_{PCE}$ . Nous avons  $\xi_{PCE} = Pr[H_{FCE}|H_{PCE}]$ . Evidemment les deux probabilités  $\beta_{FCE}$  et  $\xi_{PCE}$  dépendent du choix du seuil  $\eta$ . Nous dérivons depuis la problématique de détection introduite dans la section 2.3 les identités suivantes,

$$\beta_{FCE}(\eta) = 1 - Pr[D^* < \eta|H_{FCE}] \quad (2.15)$$

et

$$\xi_{PCE}(\eta) = Pr[D_{\lambda}^{FCE}[p] \geq \eta|H_{PCE}] \quad (2.16)$$

Dans l'équation (2.15), nous avons  $D^* = Max_{\lambda,p}\{D_{\lambda}^{FCE}[p]\}$ .

### 2.4.2 Histogrammes de la fonction de détection *READ-weld*

La figure 2.6-(a) présente l'histogramme de la *fonction de détection*  $D_{\lambda}^{FCE}[p]$ , conditionnellement à l'hypothèse  $H_{PCE}$ . La figure 2.6-(b) donne l'histogramme du maximum  $D^*$  de la *fonction de détection*, conditionnellement à l'hypothèse FCE  $H_{FCE}$ . Ces deux histogrammes ont été obtenus après analyse des données collectées depuis le pipeline décrit dans la section 2.4.3 avec *READ-weld* et  $L = 10$ .

Grâce aux fonctions de pondérations orientées circonférentiellement  $\alpha_k[n]$  (voir section 2.3.3) maximisant le contraste entre  $H_{FCE}$  et  $H_{PCE}$ , les supports des deux histogrammes présentés dans les figures 2.6-(a) et 2.6-(b) ne se superposent pas. Par conséquent, n'importe quel seuil  $\eta^*$  choisi dans l'intervalle  $[0.2, 0.45]$  mènera aux meilleures performances possibles, i.e.  $\beta_{FCE}(\eta^*) = 1$  et  $\xi_{PCE}(\eta^*) = 0$ .

### 2.4.3 Résultats de détection

Le pipeline que nous avons utilisé pour comparer *READ-weld* avec la méthode qu'utilisait TRAPIL jusque-la (nommée REF) fait 12 kilomètres de long. Le racleur qui a été utilisé pour l'inspection est un 256 sondes avec un pas d'acquisition d'1.5 millimètre. La figure 2.7 présente la probabilité de bonne détection de FCE  $\beta_{FCE}$  vis-à-vis de la probabilité d'erreur  $\xi_{PCE}$  pour *READ-weld*<sub>10</sub> (avec  $L = 10$ , i.e. la *fonction de pondération optimale*) et *READ-weld*<sub>0</sub> (avec  $L = 0$ , i.e. le *moyennage arithmétique*, voir section 2.3.3), en comparaison de la méthode REF.

Les courbes de performance présentées dans la figure 2.7 montrent que *READ-weld*<sub>10</sub> et *READ-weld*<sub>0</sub> offrent de meilleures performances tout en économisant 30% de temps d'exécution. De plus *READ-weld*<sub>10</sub> surpasse de manière significative *READ-weld*<sub>0</sub> en atteignant une probabilité de bonne détection  $\beta_{FCE}$  de 1 pour une probabilité de fausse détection bien plus faible  $\xi_{PCE}$  (inférieure à 1%) par rapport à *READ-weld*<sub>0</sub>. Cela confirme les vertues des fonctions de pondération basées sur les statistiques d'ordre élevé.

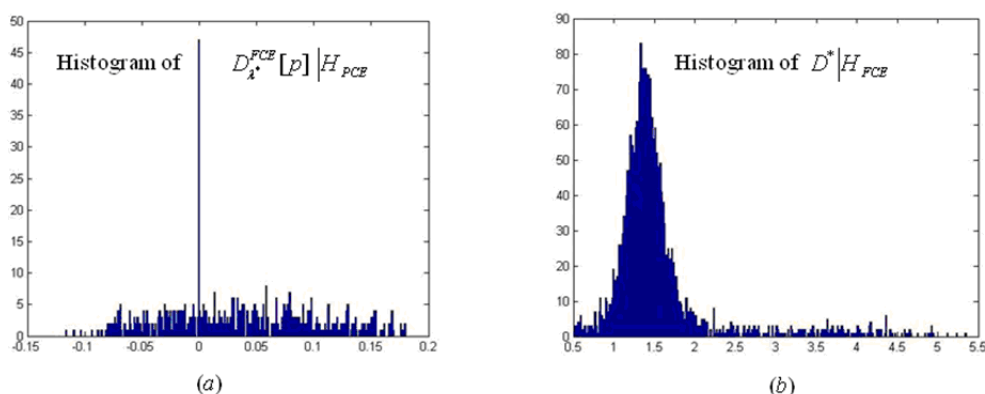


FIGURE 2.6 – (a) Histogramme de  $D_{\lambda}^{FCE}[p]$  conditionnellement à  $H_{PCE}$ , utilisant *READ-weld* avec  $L = 10$ . (b) Histogramme de  $D^*$  conditionnellement à  $H_{FCE}$  utilisant *READ-weld* avec  $L = 10$ . Les histogrammes (a) et (b) sont issus du pipeline présenté dans la section 2.4.3

## 2.5 Conclusion

Nous avons présenté dans cette partie une nouvelle méthode de détection des soudures d'aboutage contribuant à l'initiative d'inspection intelligente par racleur basée sur la technologie ultrason. *READ-weld* contient une fonction de combinaison des signaux unique basée sur les statistiques d'ordre élevées permettant à la fois une compression paramétrable des données collectées par les sondes ultrasoniques et offrant un grand contraste entre les signatures partiellement circonférentielles PCE (*comme les signatures de rides entraînées par les cintres*) et les événements entièrement circonférentiels (*comme les soudures d'aboutages*). Se reposant sur ces atouts, *READ-weld* surpasse la méthode de référence qui était utilisée par la société TRAPIL, atteignant des performances proches du 100% tout en réduisant le temps d'exécution de 30%.

Depuis la mise en service de cette méthode, TRAPIL l'a intégré totalement dans son processus d'analyse en lieu et place de la méthode REF et *READ-weld* a continué d'offrir des performances similaires aux tests sur l'ensemble des lignes qui ont été analysées depuis, ne nécessitant que très peu ou pas de corrections manuelles.

De plus, grâce à cet outil, TRAPIL dispose d'un moyen fiable et rapide lui permettant de transmettre à ses clients un carnet de tube dans des délais courts (moins de 48 heures).



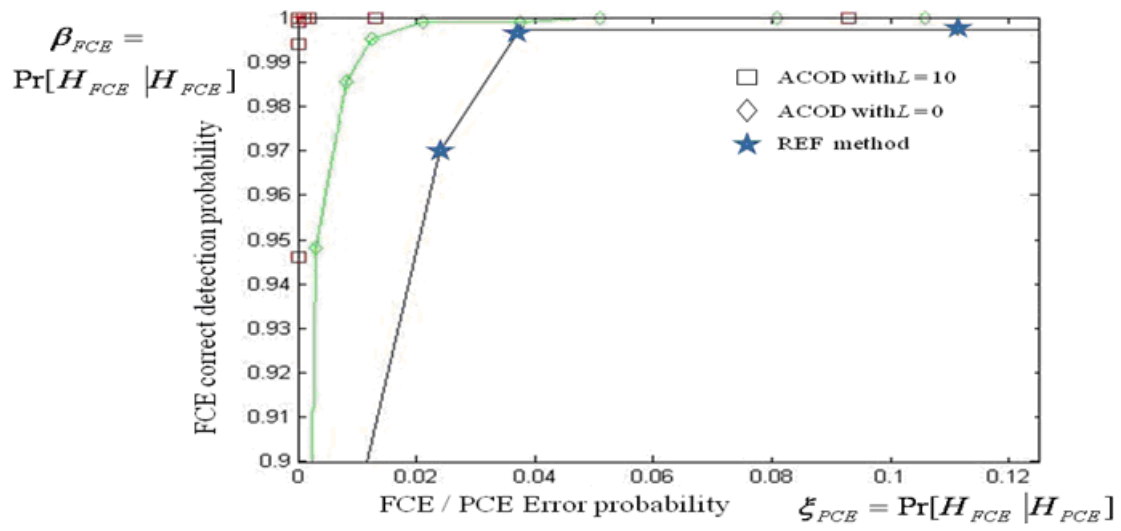


FIGURE 2.7 – Probabilité de bonne détection  $\beta_{FCE}$  vis-à-vis de la probabilité d'erreur  $\xi_{PCE}$  pour  $READ-weld_{L=10et0}$  et la méthode REF

## Chapitre 3

# Pré-détection des zones d'intérêt

### 3.1 Introduction

Une fois les soudures d'aboutage détectées, la ligne est segmentée en tubes et la détection de défauts peut débuter. Même si les tubes sont étudiés un à un, le problème de complexité reste d'actualité car la quantité de pixels à étudier au total est inchangée. C'est pourquoi la détection de défauts en elle-même est séparée en deux parties. La première partie, présentée dans ce chapitre, visera à effectuer un balayage adapté de l'ensemble du coeur de tube afin d'isoler toute zone susceptible de contenir un défaut tout en éliminant une quantité maximale de pixels dits "de fond", c'est-à-dire les parties des tubes saines. L'objectif est ici de diminuer considérablement la quantité de pixels à analyser, et de permettre de réaliser à la suite une analyse plus fine de reconnaissance (mais plus coûteuse en temps de calcul), détaillée dans le chapitre suivant.

L'objectif de la méthode de détection présente dans ce chapitre est donc double : premièrement, minimiser la quantité de pixels renvoyés, tout en maximisant le nombre de défauts contenus dans ces zones ; en particulier, un compromis en sensibilité et spécificité devra être trouvé tout en prenant en compte le contexte où on ne peut se permettre de manquer un défaut dit "critique" (susceptible d'entraîner une fuite). On définit alors dans cette partie le terme de zone d'anomalie (ZA) comme étant une partie de tube à risque, pouvant contenir un défaut, mais dont la nature est encore inconnue.

Les difficultés inhérentes à la détection de ZA sont les mêmes qu'au chapitre précédent (grande quantité de données, problème en aveugle, caractéristiques variables et fort bruit). La méthode que nous proposons, nommée  $DS^2$  (*Detection and Segmentation of Defect Signatures*), adresse à nouveau ces difficultés et est détaillée par la suite. Tout d'abord nous identifions la problématique de détection et de segmentation des signatures de défauts, puis nous présentons le détail de la méthode avant de donner les résultats obtenus au travers de notre processus expérimental entièrement détaillé.

Afin d'illustrer nos propos, nous considérons dans cette section une portion de tube, présentée dans la Figure 3.1, inspectée par un racleur 256 sondes avec un pas d'acquisition d'1.5 millimètres. Deux défauts, caractéristiques des problématiques de détection que doivent réaliser les analystes, sont présents dans cette portion, un petit délaminage et une légère corrosion à peine visible. Ce type de situation fait partie de celles les plus couramment rencontrées et montrent la difficulté au niveau psychovisuel de détecter certains défauts pouvant pourtant être considérés comme majeurs.

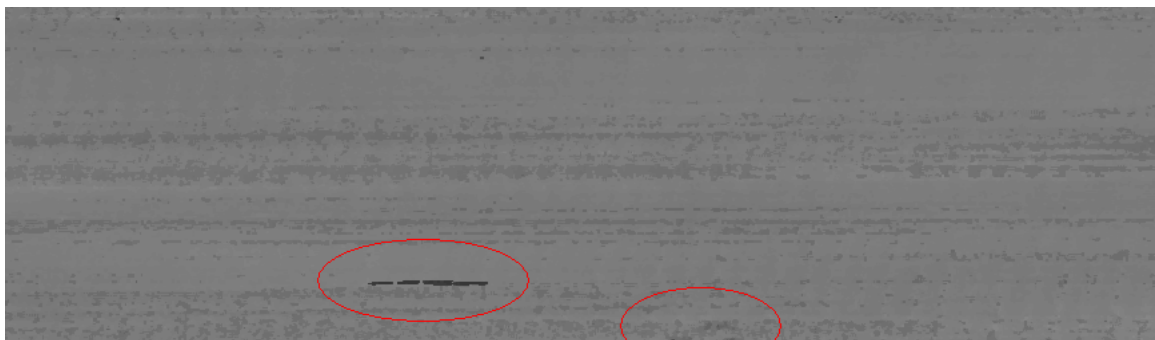


FIGURE 3.1 – Portion de tube utilisée à titre d'exemple dans ce chapitre. 256 sondes, pas d'1.5mm. Deux défauts sont présents (cerclés en rouge) : un délaminage à gauche et une légère corrosion à droite.

## 3.2 Problématique de détection des zones d'anomalie

La détection des ZA est une problématique à trois hypothèses [30], pouvant se formaliser comme suit :

$$\begin{aligned} H_0 : y_i[n] &= w_i[n] \\ H_z : y_i[n] &= x_i^z[n] + w_i[n], z \in \{B, D\} \end{aligned} \quad (3.1)$$

Dans l'équation (3.1),  $w_i[n]$  désigne un pixel appartenant à un tube sain, dont l'acquisition est propre et qui ne contient pas de défauts, alors que  $x_i^z[n]$  désigne le signal géométrique émis par la sonde  $i$  provenant de bruit ( $z = B$ ) ou d'un défaut ( $z = D$ ) localisé à l'abscisse  $n$ . Nous rappelons une fois encore que la problématique est aveugle :  $x_i^z[n]$  est inconnu et ne dispose d'aucun modèle mathématique.

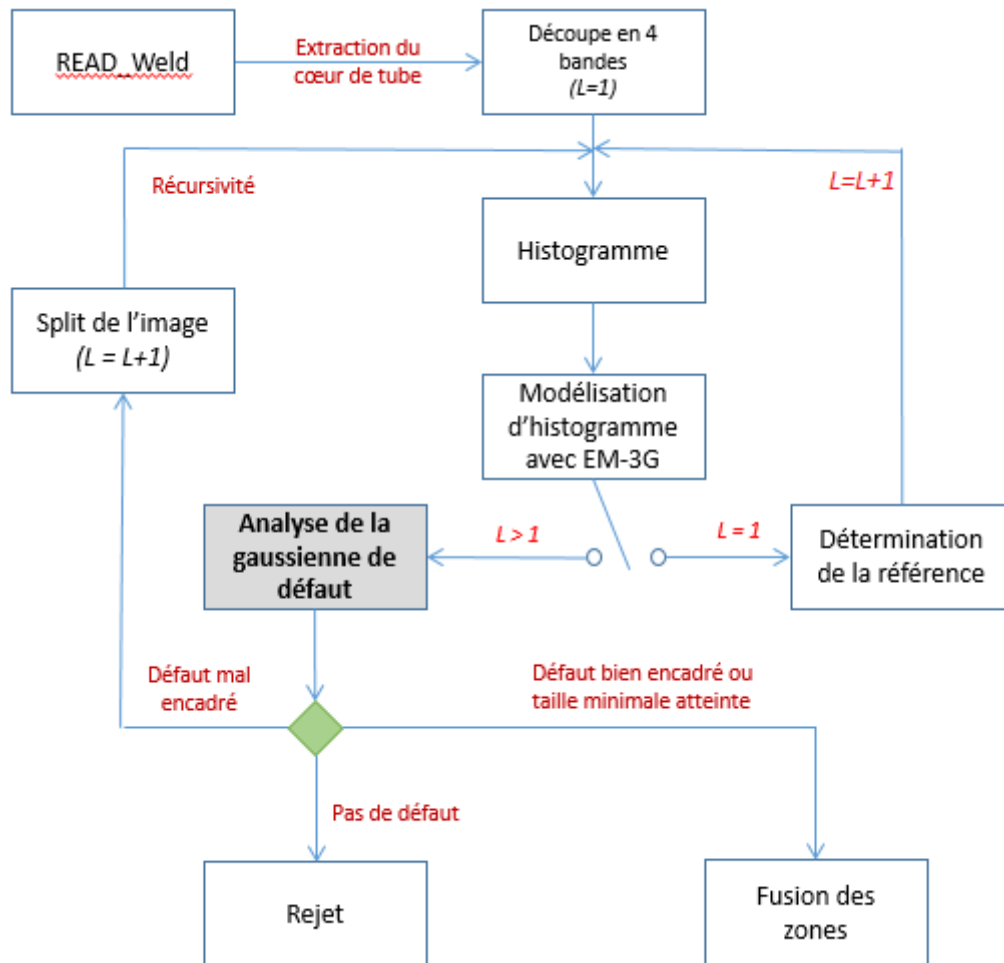
$H_0$  représente donc l'épaisseur d'un tube sain,  $H_B$  représente l'épaisseur d'une imperfection de fabrication (donc une potentielle fausse alarme), et  $H_D$  représente l'épaisseur représentant un véritable défaut ou une erreur d'acquisition ponctuelle. L'objectif de la méthode est donc de maximiser la quantité de pixels retournés appartenant à  $H_D$  tout en minimisant la quantité de pixels retournés appartenant à  $H_0$  et  $H_B$ .

## 3.3 Détection et segmentation des zones d'anomalie ( $DS^2$ )

### 3.3.1 Résumé de la méthode

Les différentes étapes de notre approche sont présentées de manière synthétique dans la Figure 3.2. La méthode  $DS^2$  fonctionne en deux temps :

- La première étape, de type prétraitement, récupère le résultat de *READ-weld* et en extrait le coeur de tube, c'est-à-dire l'intégralité du tube à laquelle nous avons retiré les ZAT au début et à la fin. Un léger débruitage de l'image à l'aide d'un filtre médian 3x3 est ensuite réalisé afin d'éliminer le bruit "poivre et sel" qui apparaît régulièrement sur les tubes. Enfin un modèle statistique "à la volée" du tube courant est construit afin de servir de référence à une étape de comparaison dans l'étape suivante.
- La seconde étape est une approche de détection récursive légèrement inspirée de la méthode *Split & Merge* : des découpes successives du tube sont effectuées, afin

FIGURE 3.2 – Aperçu de la méthode DS<sup>2</sup>

d'éliminer les zones saines au fur et à mesure et finalement de ne conserver que les ZA. Le critère de décision utilise l'algorithme *Expectation-Maximization* (EM) [18] afin de modéliser l'histogramme de la portion de tube courante à l'aide de trois gaussiennes, représentant nos trois hypothèses  $H_0$ ,  $H_B$  et  $H_D$ . Cette modélisation est alors comparée au modèle de tube sain réalisé à l'étape précédente afin de décider la présence d'une anomalie ou non. L'analyse de mixtures de gaussiennes obtenues par EM a été largement utilisée et a déjà prouvé son efficacité [4], [14], [35]. Une fois les différentes zones d'anomalies ainsi détectées, ces dernières sont isolées par une simple détection de composantes connexes.

### 3.3.2 Première étape : Prétraitements et détermination de la référence

La première étape se décompose en trois sous-étapes.

1. En premier lieu le cœur de tube est extrait. L'étude approfondie des soudures d'aboutage et des ZAT ne rentrant pas en compte dans notre problématique, elles sont ignorées. Le tube subit ensuite un médian 3x3 afin d'éliminer le bruit "poivre et sel"

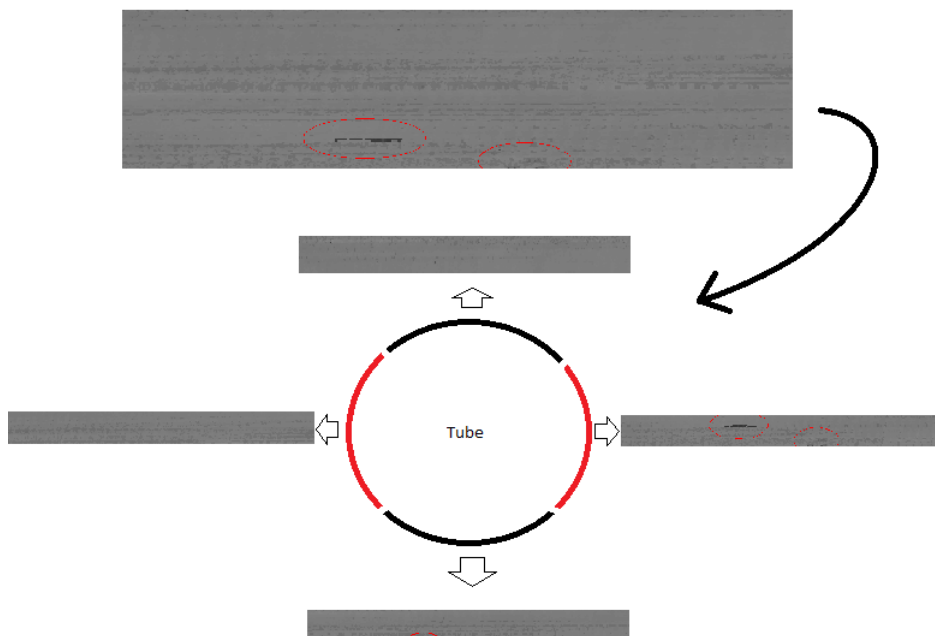


FIGURE 3.3 – Illustration de la découpe en quatre clusters de sondes.

souvent présent et très gênant pour les traitements qui vont suivre. On notera  $\tilde{y}_i[n]$  les données débruitées.

2. Ensuite,  $N_c$  clusters de  $q = \frac{N_s}{N_c}$  sondes sont construits, où  $N_s$  est le nombre de sondes du racleur utilisé. Cette découpe est réalisée car le décentrage du racleur, la plupart du temps léger et orienté vers le bas en raison de son propre poids, entraîne une légère différence sur l'acquisition circumférentielle. Typiquement, en raison de la différence de T1, due au décentrage déjà évoqué, entre les sondes situées sur la génératrice inférieure du tube et celles situées sur la génératrice supérieure, les valeurs de T2 vont être très légèrement différentes. Il s'agit d'une variation globale des valeurs d'épaisseur, très basse fréquence mais pouvant amener à des erreurs. Pour compenser cela, nous découpons donc le tube en ( $N_c = 4$ ) clusters de sondes à 45, 135, 225 et 315 degrés afin de traiter séparément le haut, le bas et les côtés du tube. La Figure 3.3 illustre la découpe en clusters de sondes.
3. Enfin un modèle de référence de la densité de probabilité de l'intensité des valeurs des pixels du tube est réalisé au moyen d'un algorithme EM appliqué au mélange de gaussiennes, initialisé avec un grand nombre  $K_{ref}$  de noyaux afin de scanner toutes les épaisseurs types que nous pouvons rencontrer dans un pipeline (typiquement de 5 à 15 millimètres). La gaussienne  $P_{ref}^n \times g_{ref}^n(x_n; \mu_{ref}^n, \sigma_{ref}^n)$  sélectionnée pour représenter les valeurs d'épaisseur saines du cluster courant  $n \in [1, N_c]$  est celle ayant la probabilité d'appartenance la plus élevée parmi tous les noyaux, s'appuyant sur l'hypothèse que l'épaisseur la plus élevée, présente en grande quantité (les sur-épaisseurs dans un tube sont rares et très ponctuelles) dans un tube entier est l'épaisseur de fond, et donc caractéristique d'un tube sain. Le fait de modéliser cette épaisseur par une gaussienne présente deux avantages : tout d'abord un échantillon de tube sain n'est jamais parfaitement homogène au millimètre près (c'est d'autant plus vrai pour

les tubes réalisés par extrusion), nous modélisons donc d'une certaine manière la texture saine d'un tube. De plus, cela nous permet d'établir une densité de probabilité de référence, caractéristique d'un fond sain et d'utiliser cette dernière directement avec les gaussiennes obtenues dans la seconde partie de l'approche proposée.

### 3.3.3 Seconde étape : Détection des zones d'anomalie

La seconde étape se fonde sur une approche récursive d'encadrement des défauts et se décompose en quatre sous-étapes :

1. La première partie consiste à modéliser de manière supervisée l'ensemble des valeurs d'épaisseur du cluster courant  $N_c$  en un mélange de trois gaussiennes non plus générale, mais propre à l'ensemble des pixels du cluster considéré. En effet, une étude approfondie des histogrammes de valeurs d'épaisseur révèle que nos trois hypothèses d'appartenance sont dans la très grande majorité des cas identifiables directement dans l'histogramme ; les parties saines de tube ( $\tilde{y}_i[n] \in H_0$ ) ont des valeurs situées logiquement autour de l'épaisseur de référence, les parties montrant des altérations de texture et du bruit de fond ( $\tilde{y}_i[n] \in H_B$ ) forment une gaussienne semblable à la précédente, mais centrée à  $\mu_B \approx \mu_0 - 1$  en millimètre. Enfin, en cas d'existence de défaut sur la partie analysée ( $\tilde{y}_i[n] \in H_D$ ), une troisième gaussienne, avec une probabilité d'appartenance bien plus faible, une moyenne inférieure à  $\mu_B$  et un écart-type bien plus grand, apparaît. Cette particularité offre ainsi la possibilité, à partir de la simple analyse de l'histogramme des valeurs d'un cluster, de déterminer l'existence ou non d'un défaut. Nous avons donc utilisé l'algorithme EM appliqué aux mélanges de gaussiennes, en l'initialisant avec  $K = 3$  noyaux tels que :

$$\begin{aligned} & (P_{H_0} = 0.45) \times g_{H_0}(x; (\mu_{H_0} = \mu_{ref}), (\sigma_{H_0} = \sigma_{ref})) \\ & (P_{H_B} = 0.45) \times g_{H_B}(x; (\mu_{H_B} = \mu_{ref} - 1), (\sigma_{H_B} = \sigma_{ref})) \\ & (P_{H_A} = 0.1) \times g_{H_A}(x; (\mu_{H_A} = \frac{\mu_{ref}}{2}), (\sigma_{H_A} = \sigma_{ref} \times 2)) \end{aligned} \quad (3.2)$$

Cette initialisation heuristique permet de coller au mieux à la problématique et donc de réduire le temps de convergence de l'EM et d'éviter de tomber dans un maximum local lors de la convergence, améliorant la qualité du résultat.

La figure 3.4 présente cette étape, partant du cluster de sondes contenant les deux défauts de notre exemple, l'histogramme correspondant et le résultat de l'EM après convergence. On peut y voir les trois gaussiennes correspondant à nos trois hypothèses, et on observe plus particulièrement le fait que la gaussienne correspondant au défaut est bien définie et donc "facilement séparable". Il est à noter qu'en cas d'absence de défaut, et lors de la convergence de l'EM, la gaussienne correspondant au défaut viendra naturellement se fondre totalement dans la gaussienne correspondant au bruit, un comportement facilement détectable, permettant de statuer directement sur la non-existence de défaut dans la portion analysée.

2. La seconde partie consiste à observer le comportement de la gaussienne de défaut après convergence dans le but de la comparer à notre gaussienne de référence obtenue

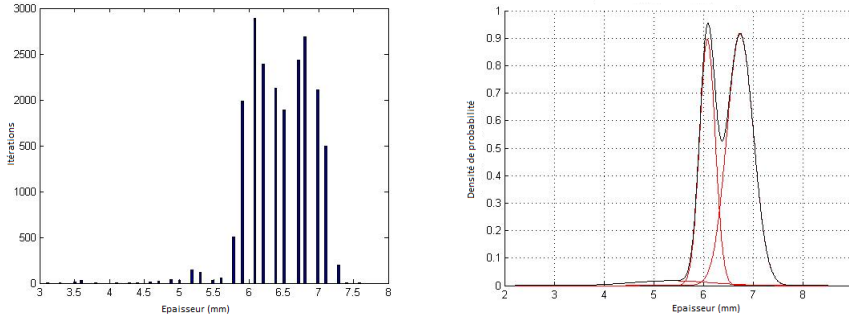


FIGURE 3.4 – Résultat de l'algorithme EM sur l'histogramme des valeurs d'épaisseurs de notre exemple illustratif.

lors de la première étape en utilisant deux critères de "classification" :

$$\begin{aligned}\Delta_\mu &= \frac{\mu_{H_D} - \mu_{ref}}{\mu_{ref}} \\ \rho_\sigma &= \frac{\sigma_{H_D}}{\sigma_{ref}}\end{aligned}\quad (3.3)$$

Dans l'équation 3.3,  $\Delta_\mu$  représente la perte de métal observée et  $\rho_\sigma$  représente la diversité des valeurs d'épaisseur présentes pour les défauts observés dans le cluster. Comme très souvent les défauts se caractérisent par, à la fois, une perte de métal élevée et une présence variée de valeurs d'épaisseur : plus ces deux valeurs sont grandes (en valeur absolue) et plus la probabilité d'observation d'un défaut réel dans le cluster est élevée.

3. La troisième partie de l'étape de détection consiste à décider ou non de l'existence d'un défaut dans le cluster observé. Les deux critères de l'équation 3.3 sont alors utilisés comme espace de classification à deux caractéristiques. Utilisant une grande population d'échantillons (environ 3000) sur diverses lignes (les critères utilisés étant tous deux relatifs par rapport à l'épaisseur de référence, nous avons pu dans cette étape utiliser des échantillons provenant de tubes et de lignes différentes) nous avons pu déterminer une séparatrice *ad hoc*. La figure 3.5 illustre la représentation de l'espace à deux caractéristiques avec la séparatrice *ad hoc*. Une approche de type régression non-linéaire nous permet alors de formaliser l'équation de la séparatrice :

$$\rho_\sigma = -\frac{(\Delta_\mu + 0.2)^{\frac{3}{2}}}{(\Delta_\mu + 0.12)}\quad (3.4)$$

L'Equation 3.4 possède deux valeurs remarquables et importantes :

- L'asymptote en  $\Delta_\mu = -0.12$ , dénotant le choix qu'une anomalie présentant une perte d'épaisseur inférieure à 12% n'est pas intéressante dans le cadre de nos travaux, qui vise à se concentrer sur les défauts les plus dangereux à l'intégrité du tube.
- La courbe croise l'abscisse en  $\Delta_\mu = -0.2$ , dénotant le choix de conserver à cette étape toute anomalie présentant une perte d'épaisseur supérieure à 20%, les informations que nous avons à cette étape n'étant pas suffisantes pour nous permettre de rejeter un défaut à la dangerosité potentielle.

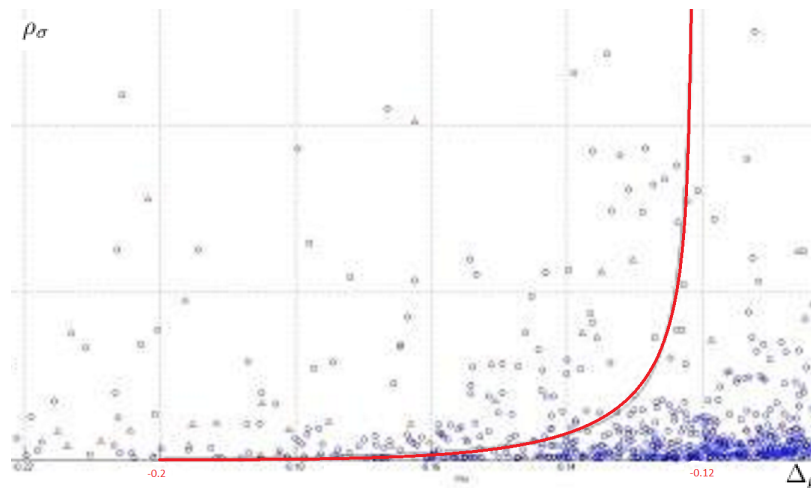


FIGURE 3.5 – Espace de représentation à deux dimensions  $(\Delta_\mu, \rho_\sigma)$ . Les défauts réels sont représentés avec des triangles, les fausses alarmes sont représentés avec des ronds. La séparatrice est en rouge.

4. Enfin, la dernière partie consiste à décider de la marche à suivre dans le déroulement de l'algorithme, en fonction des différentes informations rassemblées au-travers des étapes précédentes, tout en ajoutant un dernier critère : la probabilité d'appartenance au cluster de défaut  $P_{H_D}$ . En effet si cette probabilité est grande (typiquement supérieure à 0.66 après étude des proportions d'encadrement réalisée par les analyste) cela implique que le nombre de pixels appartenant à  $H_D$  est en grande majorité constitué des pixels appartenant à  $H_0$  ou  $H_B$ , *i.e.* le défaut est correctement encadré. De l'ensemble de cette étude, trois cas de figure peuvent se présenter :
  - Il n'y a pas de défaut dans le cluster courant : la zone entière est rejetée, passage au cluster suivant ;
  - Il y a un défaut et l'encadrement est satisfaisant : la zone observée est étiquetée comme défaut, passage au cluster suivant ;
  - Il y a un défaut mais l'encadrement n'est pas satisfaisant : la zone observée est découpée en deux (dans le sens de la longueur) et l'algorithme entre en récursivité, reprenant à la première étape sur les deux sous-portions, à la manière de l'algorithme *Split and Merge*.

Il est important de rappeler que l'épaisseur de référence  $H_0$  obtenue dans la première partie de l'approche de détection et d'encadrement est utilisée telle quelle pour l'ensemble du cluster courant et pour toutes les découpes qui en découlent, afin d'éviter un effet d'oeillères : au fil des découpes, il est possible d'en arriver à considérer une sous image totalement composée de pixels de défauts ; si la référence était alors calculée à chaque itération, les valeurs d'épaisseur du défauts seraient pris comme référence et donc la portion entière serait rejetée.

Enfin, dans le but d'accélérer l'exécution, et du fait que nous ne recherchons pas une segmentation fine du défaut par contourage au plus près, une taille minimale du cadre d'englobement a été insérée (typiquement 200 pixels de surface, ce qui correspond à environ 30 mm<sup>2</sup>) stoppant la découpe.



Au final le résultat de la méthode est un masque binaire de même taille que le tube étudié, avec des zones étiquetées 1 lorsqu'une suspicion de défaut est présente, et 0 sinon, permettant le découpage des zones d'anomalies à l'aide d'une simple segmentation binaire en composantes connexes. Cette approche, pratique d'un point de vue visualisation, permet également de refusionner naturellement des anomalies diffuses (comme une corrosion) qui auraient pu être découpées en plusieurs parties au fil de l'exécution.

### 3.4 Experimentation et Benchmarking

Comme nous l'avons dit dans les objectifs du projet, la méthode doit à la fois minimiser le nombre de pixels retournés tout en maximisant la quantité de défauts réels contenus dans les zones renvoyées. Nous avons donc déterminé trois critères d'évaluation de l'approche :

- Le pourcentage de défauts réels présents dans les zones retournées, par rapport au nombre de défauts total étiquetés par les experts, noté  $%D$
- Le nombre de zones d'anomalies retournées, donc le nombre d'itérations que l'étape de classification devra enchaîner pour une ligne donnée, noté  $nbZA$
- Le pourcentage de pixels retournés, correspondant à la somme des surfaces des zones d'anomalies par rapport au nombre de pixels d'origine de chaque tube étudié, donnant le gain de compression de notre méthode, noté  $%P$

#### 3.4.1 Protocole experimental

Les tests ont été réalisés sur une ligne sélectionnée pour ses caractéristiques représentatives : d'une longueur de 11 kilomètres, 305 défauts ont été relevés par les analystes, avec une excellente variété (43 corrosions, 129 enfoncements, 27 délaminages, 106 pertes de métal). Même si à cette étape aucune identification n'est faite sur le type de défaut, il est important de vérifier que la méthode donne de bonnes performances quel que soit le type de défaut rencontré.

La ligne possède également un changement de diamètre (de 10 à 12 pouces) et subit plusieurs changements d'épaisseur (de 7.4 à 12.5 millimètres), permettant ainsi de couvrir plusieurs cas de figure. L'outil d'inspection qui a été utilisé est équipé de 256 sondes et réalise une acquisition tous les 1.5 millimètres.

Contrairement à *READ-weld*, nous n'avons pas à notre disposition une méthode de référence éprouvée à laquelle nous comparer, mais dans le but de montrer l'efficacité de notre méthode et plus particulièrement de montrer la pertinence des choix réalisés dans le paramétrage, nous avons décidé de comparer notre séparatrice de décision *ad hoc* avec une séparatrice linéaire positionnée en  $\Delta_\mu = -0.14$ , volontairement positionnée au milieu de la zone de décision difficile modélisant de manière satisfaisante le critère décisionnel psychovisuel de l'analyste. La Figure 3.6 illustre les deux méthodes.

#### 3.4.2 Résultats

Les résultats obtenus sont présentés dans le tableau 3.1.

Les résultats sont très encourageants, puisqu'environ 90% des défauts ont été bien conservés par la méthode de détection pour un gain de compression de presque 95%. Le nombre de clusters renvoyés est grand mais raisonnable en terme de nombre d'itérations. De plus, si l'on se compare à la séparatrice linéaire, on observe une perte de 4% des défauts

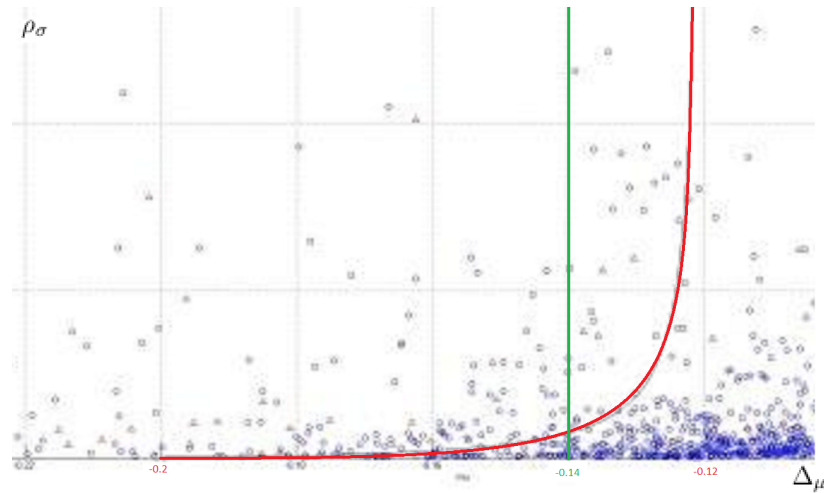


FIGURE 3.6 – Espace de représentation à deux dimensions ( $\Delta_\mu$ ,  $\rho_\sigma$ ). Les défauts réels sont représentés avec des triangles, les fausses alarmes sont représentés avec des ronds. La séparatrice ad hoc est en rouge, la séparatrice linéaire comparée est en vert.

Critère	Résultat DS <sup>2</sup>	Résultat Sép. Linéaire
%D	89.6 %	93.4 %
nbZA	≈ 52.000	≈ 186.000
%P	6.4 %	25.62 %

TABLE 3.1 – Performances des méthodes comparées sur la ligne de test

seulement mais pour une quantité de clusters renvoyés divisée par trois et surtout un gain de compression multiplié par cinq.

### 3.4.3 Illustration des résultats

Nous illustrons ici le résultat de notre méthode (Figure 3.8) et offrons une comparaison qualitative avec la méthode par seuillage linéaire (Figure 3.9) sur l'exemple de tube que nous avons utilisé comme exemple jusqu'ici (Figure 3.7).

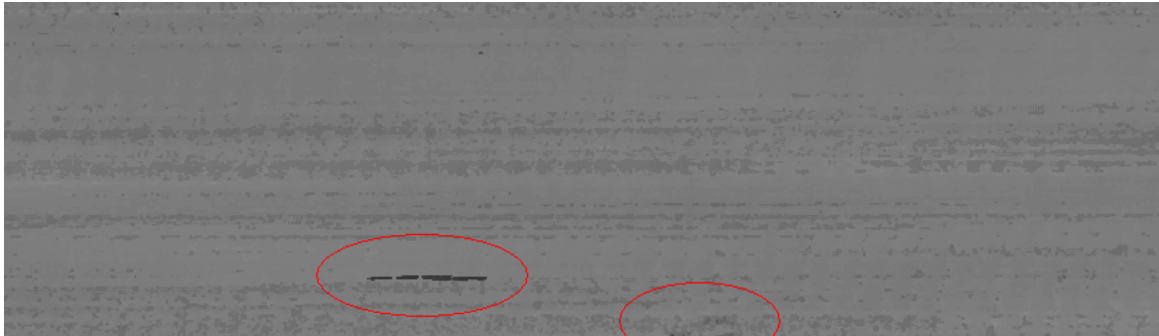


FIGURE 3.7 – *Portion de tube utilisée à titre d'exemple dans ce chapitre. 256 sondes, pas d'1.5mm. Deux défauts sont présents (cerclés en rouge) : un délaminage à gauche et une légère corrosion à droite.*

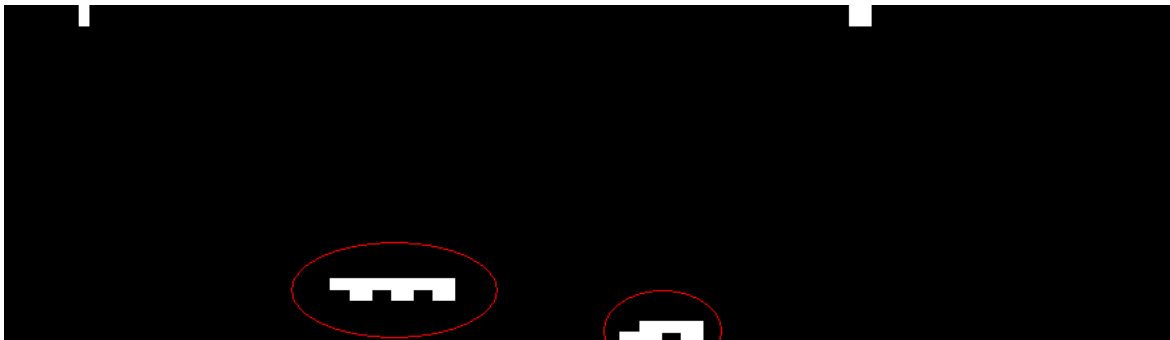


FIGURE 3.8 – *Résultat de la méthode  $DS^2$  sur l'exemple de la figure 3.7.*

Dans les deux cas les deux défauts ont bien été retournés, mais la qualité de détection est indéniablement en faveur de notre méthode.

### 3.4.4 Etude des cas d'échec

Afin de compléter l'étude précédente, nous nous sommes particulièrement intéressés aux 10% de défauts que nous avons perdus lors de nos tests, notamment pour s'assurer qu'aucun défaut dit majeur n'avait été rejeté. Il s'avère que la totalité des défauts que nous avons 'perdus' étaient des défauts identifiés comme mineurs par les analystes, présentant soit des pertes de métal infimes (inférieures à 10%), soit des défauts dont la superficie était particulièrement faible (moins de 50 pixels, soit 10 mm<sup>2</sup> environ). La figure 3.10 présente un échantillon de ces défauts en qualité d'illustration.



FIGURE 3.9 – Résultat de la méthode par seuillage linéaire sur l'exemple de la figure 3.7.

### 3.5 Conclusion

La méthode DS<sup>2</sup> apporte donc à notre problématique une réponse efficace, exploitant la modélisation d'histogramme par mixture de gaussiennes, permettant de diminuer considérablement la quantité de pixels à analyser tout en conservant une très grande majorité des défauts présents sur la ligne et en particulier ne manquant quasiment aucun défaut classifié comme majeur.

Cette compression forte de la quantité de données à analyser, couplée à une première segmentation par encadrement des anomalies retenues du même type que celle effectuée par les analystes, nous permet par la suite de nous ramener à une problématique de reconnaissance standard multi-classes, dont l'objectif est de catégoriser le type de défaut (corrosion, enfoncement, délaminage,...). Nous allons détailler cette méthode dans le chapitre suivant.

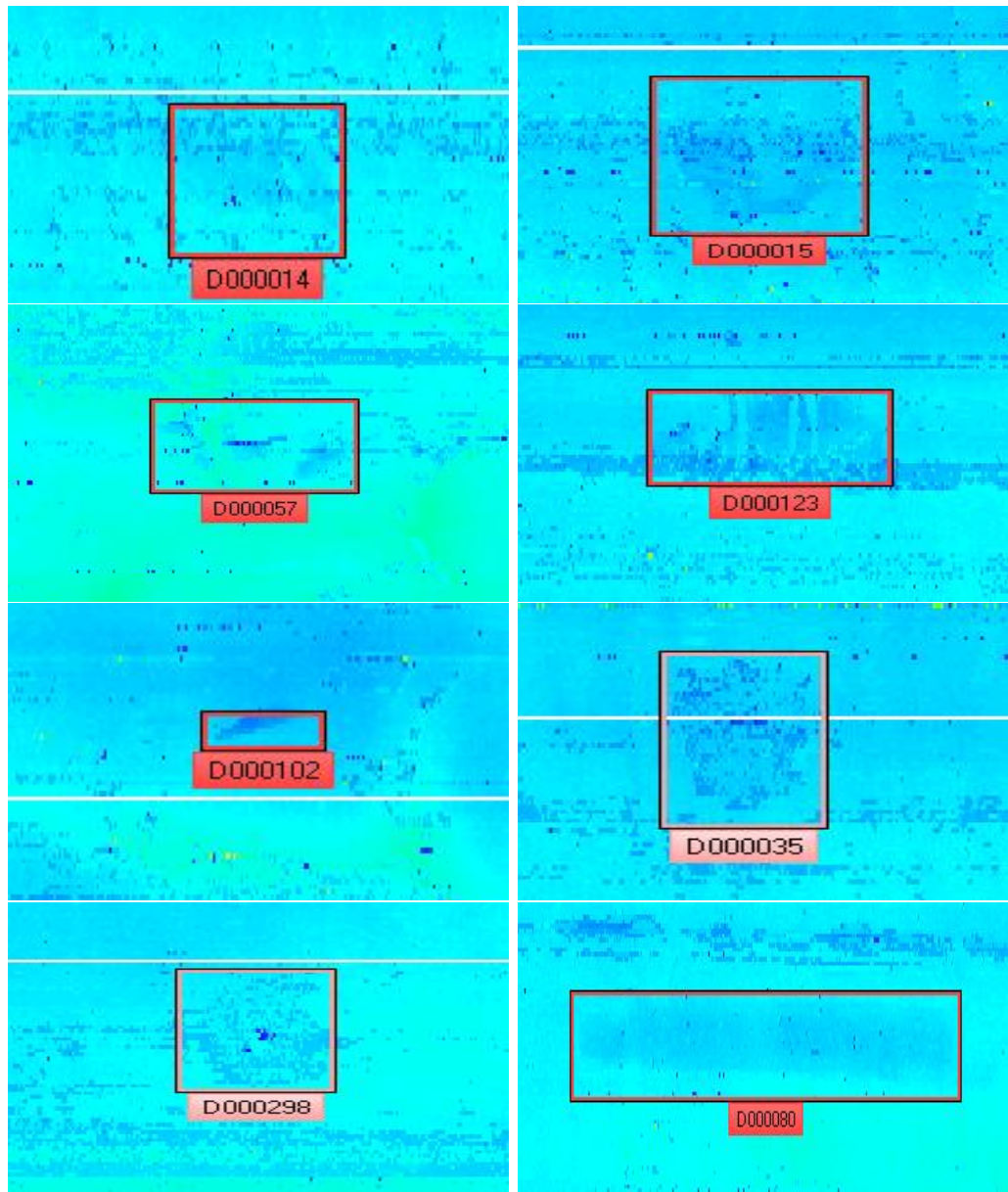


FIGURE 3.10 – Exemples de défauts "perdus" lors de l'expérimentation de la méthode DS<sup>2</sup>. Dans le sens de lecture, quatre corrosions, un délaminage, deux enfoncements et une sous-épaisseur.

## Chapitre 4

# Identification des zones d'anomalies

### 4.1 Introduction

#### 4.1.1 Application

À cette étape nous obtenons une grande quantité de sous-images que nous devons identifier. Nous revenons donc à un problème de classification standard : chaque sous image appartient alors forcément à l'une des classes suivantes :

- Enfoncement
- Sous-épaisseur
- Délaminage
- Corrosion
- Bruit

Du fait de la grande quantité d'exemples à notre disposition, au travers des analyses anti-érieures déjà réalisées par les experts, nous pouvons constituer assez facilement une base de données de défauts conséquente. De plus, grâce à l'étape de détection, les images sont pré-découpées en zone présentant potentiellement un risque et la quantité de données à traiter est donc bien moindre : contrairement aux étapes précédentes, une approche par apprentissage est alors possible. Dans ce cadre particulier, nous utilisons une approche classique d'identification en trois étapes : (i) Segmentation, (ii) Extraction de caractéristiques, et (iii) Classification.

#### 4.1.2 La reconnaissance de forme par apprentissage

##### 4.1.2.1 Principe général

Les techniques de reconnaissance de forme sont, depuis les années 1950, utilisées dans de nombreuses applications. Apparues tout d'abord pour la reconnaissance de caractères (OCR) [19], les techniques de classification se sont depuis adaptées a de nombreux domaines comme le médical, la sécurité ou plus récemment le contrôle non destructif. Leur principe est de modéliser les données que l'on souhaite identifier par un jeu de caractéristiques mathématiques. Par exemple l'une des applications didactiques de cette discipline concerne les fleurs iris. Si l'on considère qu'il existe trois familles (classes) d'iris, les iris *Setosa*, les iris *Virginica* et les iris *Versicolor* et que l'on souhaite apprendre via une machine

à les différencier, nous pouvons extraire quatre caractéristiques mathématiques simples : la longueur des pétales, la largeur des pétales, la longueur du sépale et la largeur du sépale, toutes les trois en millimètres. On nomme alors espace caractéristique l'espace à quatre dimensions, dont chaque axe correspond à une caractéristique. Dans cet espace chaque échantillon d'iris connu que nous avons à notre disposition pour réaliser l'apprentissage sera modélisé par un point dont la classe est connue (la vérité terrain). La projection de l'ensemble des points que nous avons à disposition pour réaliser l'apprentissage nous donnera alors un nuage de points appartenant à l'une des trois classes possibles. À ce moment, réaliser l'apprentissage correspond à trouver les frontières optimales dans l'espace caractéristique permettant de distinguer au mieux chaque classe à partir des données d'apprentissage. Une fois les frontières définies, l'apprentissage est terminé : tout nouvel iris présenté dont la classe est inconnue pourra être identifié (ou classifié) en le projetant dans notre espace de caractéristiques et en lui affectant la classe de la "zone" dans laquelle il est identifié. La Figure 4.1 illustre le concept de classification avec un problème simple, différencier les images de roses du reste.

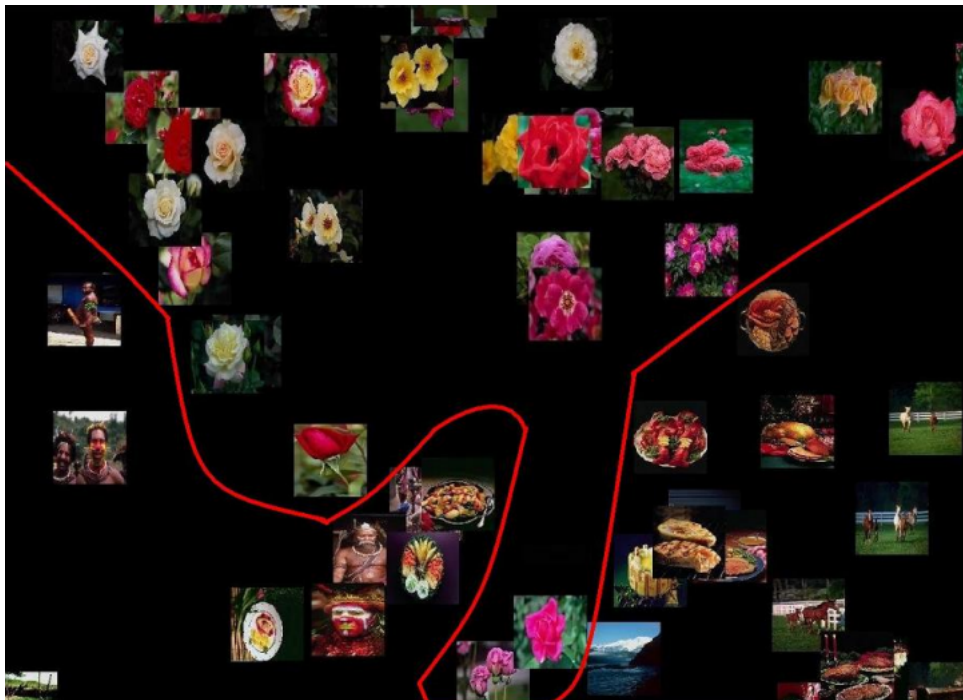


FIGURE 4.1 – *Illustration didactique d'un problème de classification où l'on cherche à séparer les images de rose des autres images.*

Bien entendu, le choix des caractéristiques est primordial : plus les caractéristiques permettent une séparation claire des nuages de points dans l'espace (on parle alors de caractéristiques possédant un fort pouvoir discriminant) plus il sera aisé de trouver une frontière simple permettant d'affecter les classes. Dans le cas idéal, un simple plan dans l'espace permettra de séparer deux classes, on parlera alors de classes linéairement séparables. Mais bien souvent même avec de bonnes caractéristiques on observera un mélange dans les nuages de points de deux classes et il sera impossible de les séparer avec un simple plan : on parle alors de recouvrement. La figure 4.2 est une représentation des nuages de points de la base de donnée Iris en ne considérant que deux caractéristiques :

la longueur du sépale et la longueur des pétales. On peut voir que les iris Setosa sont linéairement séparables des deux autres. Cependant les classes Virginica et Versicolor ont un recouvrement.

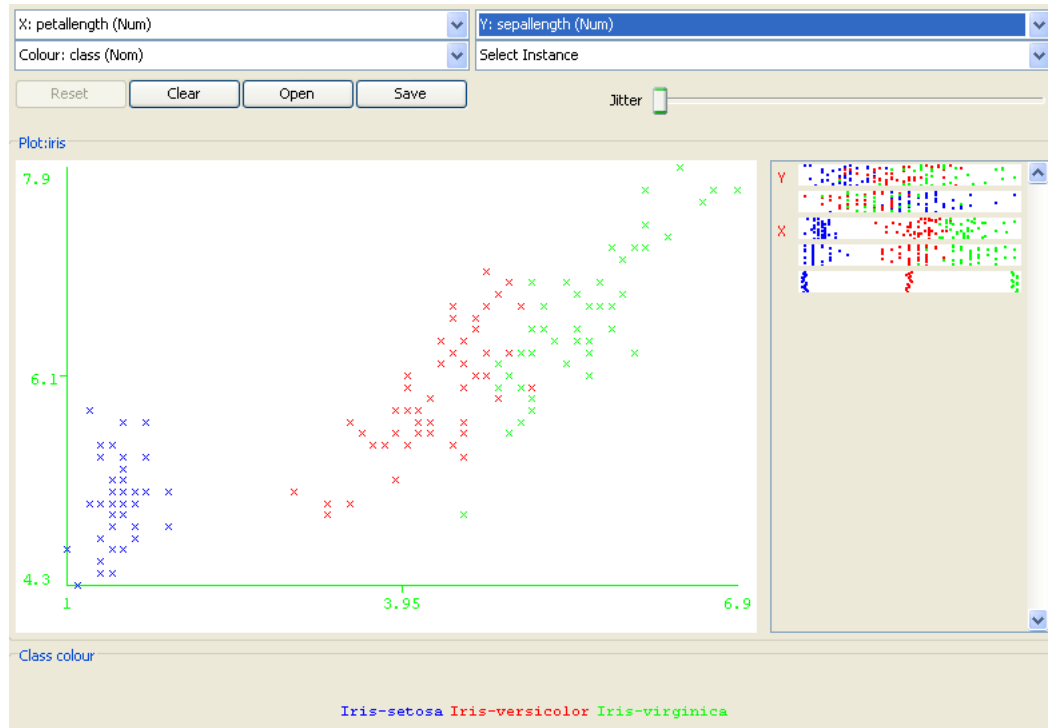


FIGURE 4.2 – Nuage de points sur la base de donnée Iris avec deux caractéristiques : longueur du sépale et la longueur des pétales.

#### 4.1.2.2 Le classifieur

Un classifieur est un algorithme qui va permettre de trouver de façon automatique la (ou les) frontières permettant de séparer au mieux deux (ou plus) classes. Selon leur fonctionnement, les temps d'apprentissage (temps demandé afin de fixer les frontières associées à un nouveau problème) et les temps de décision (temps demandé pour donner une décision, une fois l'apprentissage réalisé, sur un nouvel échantillon dont la classe est inconnue) peuvent être variables, mais il s'agit généralement d'un compromis : dans la plupart des cas un classifieur apprenant très vite mettra plus de temps à donner une décision qu'un classifieur qui demande beaucoup de temps pour apprendre.

Le classifieur se caractérise également par ses performances ; déterminer les performances d'un classifieur est assez naïf : il s'agit simplement des probabilités de bonne décision sur un jeu de tests différent du jeu d'apprentissage. Cependant, comme dans beaucoup d'applications confondre certaines classes entre elles peut être moins grave que d'en confondre d'autres, une matrice de confusion est alors le moyen le plus efficace pour juger des performances d'un classifieur dans le cas d'un problème à plus de deux classes (ce qui est notre cas).

Au-delà des performances, les classifieurs sont également caractérisés selon leur comportement vis-à-vis de certaines notions importantes : on parle de classifieur instable lors-



qu'une légère modification du jeu d'apprentissage peut entraîner une grande différence dans les frontières résultantes après apprentissage. On dit aussi qu'un classifieur à tendance à sur-apprendre lorsque celui-ci crée des frontières extrêmement fines et précises dans le but de séparer totalement les classes lors de l'apprentissage et qu'il devient alors peu généralisable à de nouveaux exemples. Ces comportements ne sont pas forcément directement impactants sur les performances. Néanmoins, selon l'application, on peut être amené à les rechercher ou à vouloir les éviter.

Il existe de nombreuses familles de classifieurs, des plus simples, comme le K-plus proche voisin (Kppv) qui se "contente" de mesurer la distance dans l'espace caractéristique entre un nouvel échantillon et les échantillons d'apprentissage, décidant de la classe la plus représentée parmi les K plus proches, aux classifieurs plus complexes comme les réseaux de neurones [20], ou encore les *Support Vector Machine* (SVM) [27]. Il est également usuel de mettre en relation plusieurs classifieurs simples pour améliorer leurs performances, soit les uns à la suite des autres (cascade de classifieurs [31]), soit en mettant en parallèle un grand nombre de classifieurs élémentaires comme c'est le cas pour les Forêts aléatoires [2] ou le Boosting [8].

### 4.1.2.3 Les Forêts aléatoires

Le principe des Forêts Aléatoires est d'utiliser les particularités d'un classifieur très simple, l'arbre de décision binaire [3]. En effet, l'arbre de décision est un classifieur rapide mais extrêmement instable : réalisant des coupes successives et optimales de l'espace de caractéristiques pour tracer ses frontières (*cf* illustration de la figure 4.3), la moindre petite modification du jeu d'apprentissage entraîne alors une structure d'arbre résultante potentiellement totalement différente.

L'idée des Forêts aléatoires est alors de mettre en parallèle un grand nombre d'arbres de décision, tout en injectant une petite portion d'aléatoire dans le jeu d'apprentissage de chaque arbre unitaire. Les structures résultantes sont alors bien différentes et la mise en relation de tous ces arbres par vote à la majorité permet d'obtenir un léger effet de flou au niveau des frontières du classifieur dans sa globalité, améliorant grandement ses capacités de généralisation. De nombreuses façons d'injecter de l'aléatoire dans le processus décisionnel de l'arbre binaire ont été proposées, offrant autant de variantes de forêts aléatoires, mais la plus utilisée reste la variante proposée par Breiman (*Random Forest - Random Input*) dans son article fondateur [2], qui consiste à injecter de l'aléatoire de deux manières :

- Le Bagging (*Bootstrap Aggregating*) consiste à créer un jeu d'apprentissage unique pour chaque classifieur unitaire en réalisant un tirage aléatoire avec remise des échantillons de la base d'origine, sans changer le nombre d'échantillons. De ce fait certains échantillons vont se trouver en plusieurs exemplaires, d'autres seront absents, cela pour chaque classifieur unitaire.
- Le RFS (*Random Feature Selection*) consiste, lors du développement récursif de l'arbre binaire, à ne plus sélectionner la caractéristique optimale pour réaliser la découpe selon l'ensemble des caractéristiques existantes, mais à choisir cette caractéristique parmi un sous-ensemble tiré aléatoirement sans remise. La découpe est alors modifiée, entraînant une structure résultante souvent très différente de l'optimal.

La Forêt aléatoire résultante est un classifieur aux performances très intéressantes, extrêmement rapide en décision, possédant une excellente capacité de généralisation et fonc-

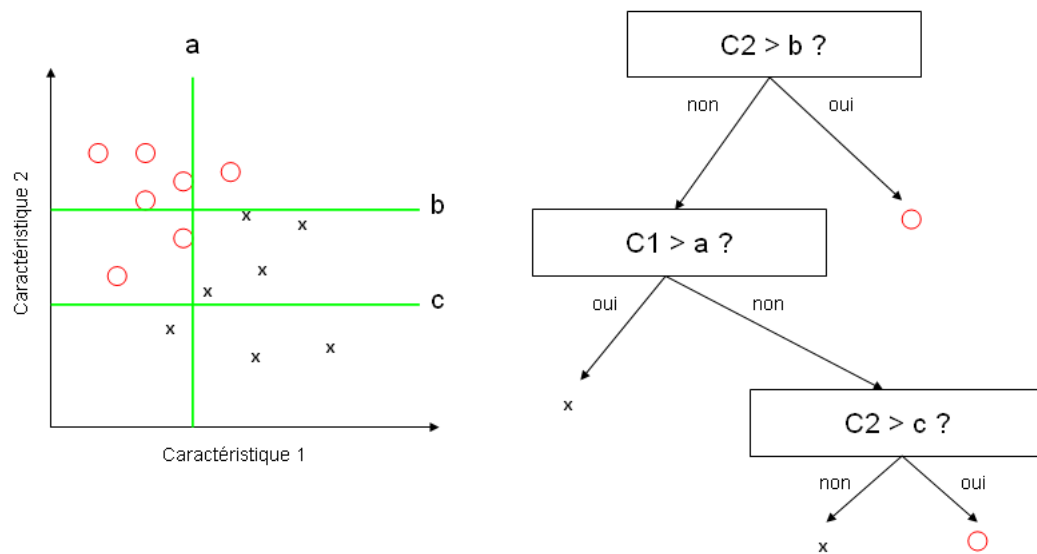


FIGURE 4.3 – Exemple de déroulement de l’algorithme d’un arbre de décision binaire sur un cas simple à deux classes et deux caractéristiques .

tionnant très bien pour de grandes dimensionnalités de l’espace caractéristiques (l’arbre de décision réalisant une succession de découpes unidimensionnelles, la quantité de caractéristiques n’impacte pas les performances). C’est pour ces raisons que notre choix premier s’est porté sur ce classifieur, semblant parfaitement adapté à notre problématique. Concernant les SVM et les réseaux de neurones, qui seront également utilisés à titre de comparaison de performances, nous renvoyons le lecteur à respectivement [27] et [20] pour une description complète.

## 4.2 Problème d’identification des zones d’anomalies

Si nous reprenons les hypothèses du chapitre précédent (voir Section 3.2), l’hypothèse  $H_0$  n’a plus cours ici : les sections entièrement saines ne peuvent avoir été sélectionnées. Nous nous situons donc uniquement dans l’hypothèse  $H_z$ , toujours avec  $z \in B, A$ . Cependant l’hypothèse  $H_A$  est ici décomposée selon le type de défaut en présence. L’objectif de la méthode est donc double :

- Premièrement réussir à différencier les défauts du bruit (fausses alarmes de l’étape précédente) ;
- Deuxièmement différencier les classes de défauts.

Il est à noter que le premier objectif est plus important que le second : dans le cas pratique, un défaut confondu avec un autre défaut pourra facilement être réétiqueté lors de la vérification faite par l’analyste, alors qu’un défaut étiqueté comme fausse alarme sera perdu.

## 4.3 Identification des zones d'anomalies

### 4.3.1 Aperçu de la méthode

Notre approche à ce stade est assez classique dans un cadre de classification par apprentissage : nous disposons d'une base de données d'images de défauts et de fausses alarmes (obtenues à l'étape précédente, et après étiquetage manuel de l'ensemble des retours) et devons apprendre un classifieur permettant de différencier directement les différents types de défauts des fausses alarmes.

Pour ce faire nous réalisons tout d'abord une première étape de segmentation fine afin de bien encercler le défaut, puis réalisons une extraction de caractéristiques. Une fois les vecteurs de caractéristiques obtenus, l'apprentissage s'ensuit naturellement.

### 4.3.2 Première étape : Segmentation fine

La segmentation par encadrement de l'étape précédente était suffisante pour détecter les défauts dans le tube, mais dans le but de faciliter l'extraction de caractéristiques nous avons besoin d'obtenir des contours précis du défaut. Cela est plus aisé car la taille des images étudiée étant dorénavant petite, le nombre de modes présents est alors réduit à deux : en effet une même sous-image contient alors systématiquement des pixels appartenant à  $H_0$ , c'est-à-dire des parties de tube sain, provenant de la segmentation imparfaite, qui vont se trouver autour du défaut ou de la fausse alarme détectée, et les pixels appartenant soit à  $H_A$ , soit à  $H_B$ . Evidemment dans le cas concret où un défaut serait recouvert par du bruit, la possibilité d'un triple mode existe, mais dans ce cas la discrimination sera de toute manière impossible par manque d'information fiable ; c'est pourquoi nous nous permettons de faire cette approximation.

La problématique de segmentation devient alors très simple, il nous faut différencier deux modes en présence au moyen d'une méthode rapide et précise : la binarisation par seuillage automatique d'Otsu s'impose alors d'elle même. Cette méthode robuste a déjà montré son efficacité par le passé [23].

Assez logiquement, cette binarisation permet d'obtenir des résultats tout à fait satisfaisants sur l'ensemble des tests que nous avons réalisés, comme on peut le voir Figure 4.4. On peut en particulier noter que les défauts sont segmentés assez précisément, facilitant l'extraction de caractéristiques qui intervient ensuite. Sur ces images binarisées va également être réalisée une détection de composantes connexes afin de ne prendre en compte que la composante présente la plus grande pour les calculs d'extraction de caractéristiques.

### 4.3.3 Seconde étape : Extraction de caractéristiques

Une fois la segmentation fine obtenue, nous pouvons extraire le vecteur de caractéristiques permettant l'apprentissage du classifieur. La problématique traitée dans le cadre de ce travail étant nouvelle dans le domaine de la reconnaissance de forme nous n'avons pas d'*a priori* en ce qui concerne les caractéristiques à extraire. Nous avons donc choisi de procéder en deux temps : en premier nous avons extrait toutes les caractéristiques qui nous semblaient pertinentes, puis en second nous avons réalisé une analyse en composante principale (ACP) [34] nous permettant d'identifier les caractéristiques les plus pertinentes et les plus discriminantes.

Nous avons exploré trois familles de caractéristiques :

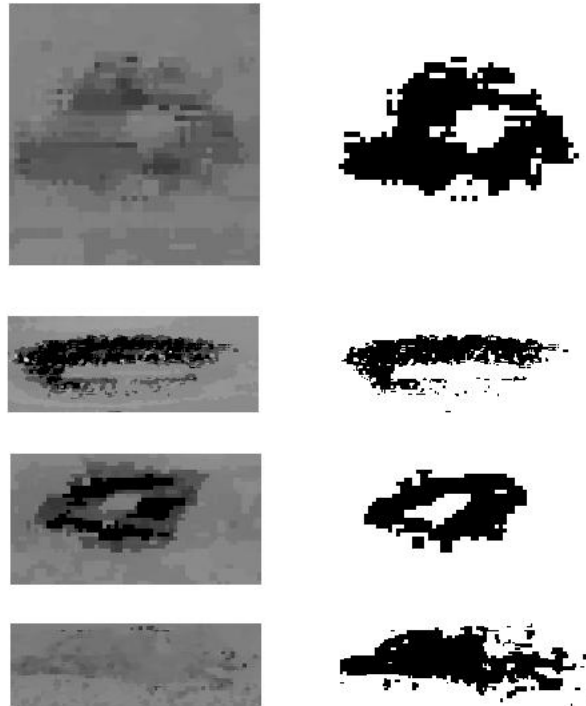


FIGURE 4.4 – Exemples de résultats de binarisation sur des extractions de défauts réalisées au chapitre précédent.

- Les caractéristiques orientées contexte, dédiées à notre problématique
- Les caractéristiques généralistes liées à la forme du défaut
- Les caractéristiques généralistes liées à la texture du défaut

La première famille de caractéristiques contient les caractéristiques orientées contexte, dédiée à notre problématique . On trouve ici :

- La largeur de l'image
- La hauteur de l'image
- Le nombre de composantes connexes dans l'image
- La surface de la plus grande composante connexe
- La perte de métal relative maximale au sein de la plus grande composante connexe
- L'écart-type des valeurs d'épaisseur au sein de la plus grande composante connexe
- Le gradient moyen sur les contours de la plus grande composante connexe

La seconde famille comporte des caractéristiques dédiées à représenter la forme de notre défaut. Ce sont des caractéristiques connues, utilisées dans bon nombre d'applications ; on peut les retrouver notamment dans la fonction *regionprops* de MATLAB. Nous prenons en compte pour ces calculs uniquement la plus grande composante connexe de l'image :

- L'excentricité de la forme
- Le diamètre équivalent d'un cercle de même aire que la forme

- le ratio de surface entre la forme et l'image
- La longueur du grand axe de l'ellipse collant au plus près de la forme
- La longueur du petit axe de l'ellipse collant au plus près de la forme
- L'orientation en degrés de l'ellipse collant au plus près de la forme
- Le périmètre de la forme
- La compacité de la forme (aire de la forme divisée par l'aire convexe)

Enfin, la troisième famille comporte des caractéristiques dédiées à représenter la texture de notre défaut. Pour ce faire nous avons utilisé les matrices de co-occurrence (*Gray Level Co-occurrence Matrix, GLCM*).

La matrice de co-occurrence  $\mathcal{MC}(d, \theta)$  est définie pour un couple  $(a, b)$  de niveaux de gris et une translation  $t$ . La translation  $t$  est déterminée par le couple  $(d, \theta)$  où  $d$  est la longueur de  $t$  et  $\theta$  l'angle formé par  $t$  et l'horizontale.  $\mathcal{MC}(d, \theta)(a, b)$  est donc le nombre de couples de sites  $(s, s+t)$  de la région considérée, séparés par le vecteur de translation  $t$  et tels que  $s$  ait pour niveau de gris  $a$  et  $s+t$  pour niveau de gris  $b$ .

Ainsi, pour une image codée sur  $N$  niveaux de gris, la matrice de co-occurrence sera de taille  $N \times N$ . L'exemple ci-dessous est réalisé pour  $t = (0, -1)$  :

$$\text{image source} = \begin{pmatrix} 1 & 2 & 1 & 3 & 4 \\ 2 & 3 & 1 & 2 & 4 \\ 3 & 3 & 2 & 1 & 1 \end{pmatrix} \quad \text{matrice de co-occurrence} = \begin{pmatrix} \begin{matrix} a \backslash b \\ 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{matrix} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 2 & 0 & 0 & 0 & 0 \\ 2 & 0 & 1 & 0 & 2 & 0 & 0 & 0 \\ 3 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 4 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 6 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 7 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

En pratique, la matrice de co-occurrence est calculée pour des distances entières allant de 1 à 5 et pour les angles  $0^\circ, 45^\circ, 90^\circ, 135^\circ$  et  $180^\circ$ , compte-tenu des propriétés de symétrie suivantes :

$$\begin{aligned} \mathcal{MC}(d, 0^\circ) &= \mathcal{MC}^T(d, 180^\circ) \\ \mathcal{MC}(d, 45^\circ) &= \mathcal{MC}^T(d, 225^\circ) \\ \mathcal{MC}(d, 90^\circ) &= \mathcal{MC}^T(d, 270^\circ) \\ \mathcal{MC}(d, 135^\circ) &= \mathcal{MC}^T(d, 315^\circ) \end{aligned}$$

TABLE 4.1 – Propriétés de symétries des matrices de co-occurrence

Il est possible d'extraire de cette matrice un grand nombre de paramètres. Néanmoins, certains sont plus parlants que d'autres, et pour une matrice  $\mathcal{MC}$  de co-occurrence donnée, 17 paramètres sont couramment calculés [9], [29], [6]. Nous donnons ci-dessous quelques exemples illustratifs.

Soit  $K$  le nombre de total de translation  $t'$  de niveaux de gris dans l'image originale. Le second moment angulaire (ou énergie ou homogénéité) mesure l'homogénéité du défaut.

Sa valeur est d'autant plus faible qu'il y a peu de zones homogènes :

$$AngScMom = \frac{1}{K^2} \sum_{a,b} \mathcal{MC}(a,b)^2 \quad . \quad (4.1)$$

Le contraste mesure les variations locales des niveaux de gris :

$$Contrast = \frac{1}{K} \sum_{k=0}^{N-1} k^2 \sum_{|a-b|=k} \mathcal{MC}(a,b) \quad . \quad (4.2)$$

Le moment différentiel inverse a un comportement inverse du contraste :

$$InvDfMom = \frac{1}{K} \sum_{a,b} \frac{1}{1 + (a-b)^2} \mathcal{MC}(a,b) \quad . \quad (4.3)$$

L'entropie fournit un indicateur du désordre de la texture :

$$Entropy = -\frac{1}{K} \sum_{a,b} \mathcal{MC}(a,b) \log \left( \frac{\mathcal{MC}(a,b)}{K} \right) \quad . \quad (4.4)$$

Plus globalement, nous avons extrait des matrices de co-occurrence les caractéristiques suivantes :

- L'auto-correlation
- La corrélation
- La prééminence des cluster
- L'ombre des cluster
- La dissimilarité
- L'entropie
- La variance
- La moyenne des sommes
- La variance des sommes
- L'entropie des sommes
- La variance des différences
- L'entropie des différences
- L'information des mesures de corrélation 1
- L'information des mesures de corrélation 2
- La différence inverse normalisée
- L'énergie
- Le contraste
- L'homogénéité
- L'homogénéité des différences inverses
- Le maximum de probabilité

Ces trois familles mises ensemble nous donnent un vecteur de 35 caractéristiques. N'ayant pas *a priori* sur la quantité de caractéristiques à conserver (le travail de classification à haute dimensionnalité n'est pas forcément gênant, il réduit juste le choix des classifieurs utilisables), et du fait que nous recherchons les meilleures performances possibles à cette étape, nous réalisons une ACP dans le but de supprimer uniquement les

caractéristiques superflues, donc pauvre en information. Nous avons donc choisi la combinaison de caractéristiques offrant 99% du pouvoir discriminant, ce qui a permis d'identifier et de rejeter les caractéristiques suivantes :

- Le diamètre équivalent d'un cercle de même aire que la forme
- Le contraste des GLCM
- L'énergie des GLCM
- L'homogénéité des GLCM
- Le maximum des probabilité des GLCM
- L'homogénéité des différences inverses GLCM
- La variance des différences GLCM

Nous obtenons donc un vecteur final de 28 caractéristiques, possédant une faible corrélation et nécessitant peu de temps de calcul pour être compilées. Ces caractéristiques peuvent donc être utilisées pour réaliser l'apprentissage d'un classifieur.

#### 4.3.4 Troisième étape : Apprentissage et Classification

Une fois le vecteur de caractéristiques défini, il est alors nécessaire de construire une base de données afin de réaliser l'apprentissage et le test d'un classifieur. Nous avons donc naturellement choisi d'utiliser les résultats du chapitre précédent afin de concevoir cette base. Un problème est cependant apparu : le déséquilibre de la base d'apprentissage. En effet, l'étape précédente nous a fourni 52.000 échantillons de fausses alarmes pour environ 350 défauts, entre 50 et 100 dans chaque classe (corrosion, délaminage, sous-épaisseur, enfoncement). Nous avons donc un fort déséquilibre en terme de quantité d'échantillons, ainsi qu'une population de défauts finalement faible selon les classes. Un tel déséquilibre est handicapant lors de l'apprentissage d'un classifieur car cela peut entraîner un sur-apprentissage des fausses alarmes, de sorte que si nous apprenons un classifieur sur une telle base de données, ce dernier se contentera de répondre en fausse alarme systématiquement.

Il nous est de plus impossible dans le temps imparti par la thèse de rassembler plus de données de défauts ; en effet les méthodes ont été testées sous MATLAB, augmentant considérablement le temps de calcul et nous empêchant donc de compiler d'autres lignes de test. Cette difficulté est néanmoins assez classique dans les problématiques de reconnaissance de forme, surtout dans le médical où il est souvent bien plus aisé d'obtenir des images de patients sains que des images de patient présentant la pathologie considérée. Nous avons donc imaginé deux solutions pour rééquilibrer la base d'apprentissage :

- Diminuer simplement le nombre de fausses alarmes (en réalisant un tirage aléatoire sans remise) ;
- Augmenter artificiellement le nombre de défauts en dupliquant les échantillons disponibles en ajoutant un léger bruit gaussien sur chaque caractéristique

La première solution est bien entendu plus justifiable car cela n'implique pas de générer artificiellement des données qui n'existent pas. Cependant cela entraîne une destruction complète d'un très grand nombre de données que nous avons à notre disposition sur le comportement des fausses alarmes. De plus dans le cas pratique, la quantité de fausses alarmes que nous avons à traiter est généralement bien plus importante que le nombre de défauts réels. Nous avons donc choisi de conserver les deux possibilités et de les tester.

Reste alors le choix du classifieur. La seule restriction que notre application impose quant à ce choix est d'avoir un temps de décision rapide. C'est pourquoi nous avons

	Fausse alarme	Corrosion	Perte de métal	Enfoncement	Délaminage
Fausse alarme	<b>493</b>	11	14	12	35
Corrosion	24	<b>8</b>	11	3	3
Perte de métal	26	5	<b>45</b>	7	2
Enfoncement	6	5	3	<b>75</b>	3
Délaminage	38	0	2	8	<b>124</b>

TABLE 4.2 – *Matrice de confusion résultat pour le PMC sur la base de tests avec suppression de fausses alarmes*

selectionné trois classifieurs réputés pour leurs bonnes performances et possédant un temps de décision rapide : Les Forêts aléatoires, le réseau de neurones (ou Perceptron Multi-Couches, PMC) et les *Support Vector Machine* (SVM).

## 4.4 Performances

Nous avons utilisé les résultats obtenus au travers du chapitre précédent pour tester ces différents classifieurs. Chaque échantillon retourné par notre méthode DS<sup>2</sup> a été étiqueté à la main et nous avons benchmarké nos trois classifieurs et les deux solutions de rééquilibrage, selon le protocole de test en *cross-validation 10-fold* : nous tirons 90% de nos échantillons aléatoirement pour apprendre et testons sur les 10% restants, répétant cette manoeuvre 10 fois et en moyennnant les résultats, afin de lisser les performances et éviter de tomber dans des cas particuliers dus aux tirages qui ne seraient pas représentatifs.

Nous avons utilisé le logiciel Weka [22] pour tester les différents classifieurs sur notre base de donnée. Les trois classifieurs ont été utilisés avec un paramétrage standard : le réseau de neurone possède (nombre de classe + nombre d'attributs) couches cachées, les Forêts aléatoires ne subissent pas d'élagage, et le SVM utilise un noyau (kernel) de base radiale d'équation type  $\exp^{-\gamma*(u-v)^2}$  avec  $\gamma = -\frac{1}{2\sigma^2}$  ( $\sigma$  hyperparamètre), et  $u$  et  $v$  deux échantillons.

Les résultats de classification obtenus sont présentés sous la forme d'une matrice de confusion : les lignes représentent la classe d'appartenance, les colonnes représentent la classe qui a été affectée par le classifieur, en nombre d'échantillons. De ce fait les échantillons situés sur la diagonale haut-gauche / bas-droit sont les échantillons qui ont bien été classifiés.

Les tableaux de 4.2 à 4.7 présentent les résultats obtenus sur la base de tests pour nos trois classifieurs, et selon nos deux protocoles de rééquilibrage de la base : suppression de fausses alarmes ou ajout artificiel de défauts.

Comme nous nous y attendions, le SVM ne donne pas de bons résultats : la modélisation par hypersphère déformée n'est pas adaptée à notre problématique car le nombre de caractéristiques est important et surtout des défauts appartenant à la même classe peuvent être morphologiquement très différents ; les clusters de nos classes dans l'espace caractéristiques sont donc probablement bien loin d'une hypersphère. Cependant, les réseaux de neurones et les forêts aléatoires offrent des performances très intéressantes : même en diminuant drastiquement le nombre de fausses alarmes pour équilibrer la base nous obtenons des probabilités de bonne détection d'environ 80% pour les forêts aléatoires



	Fausse alarme	Corrosion	Perte de métal	Enfoncement	Délaminage
Fausse alarme	<b>51619</b>	15	99	84	178
Corrosion	546	<b>3508</b>	267	389	239
Perte de métal	509	125	<b>7586</b>	239	126
Enfoncement	246	145	115	<b>8490</b>	296
Délaminage	875	97	132	312	<b>15956</b>

TABLE 4.3 – Matrice de confusion résultat pour le PMC sur la base de tests avec augmentation artificielle du nombre de défauts

	Fausse alarme	Corrosion	Perte de métal	Enfoncement	Délaminage
Fausse alarme	<b>557</b>	0	0	0	8
Corrosion	49	<b>0</b>	0	0	0
Perte de métal	85	0	<b>0</b>	0	0
Enfoncement	91	0	0	<b>0</b>	1
Délaminage	120	0	0	0	<b>52</b>

TABLE 4.4 – Matrice de confusion résultat pour le SVM sur la base de tests avec suppression de fausses alarmes

	Fausse alarme	Corrosion	Perte de métal	Enfoncement	Délaminage
Fausse alarme	<b>48394</b>	4	330	854	2413
Corrosion	3958	<b>5</b>	242	639	105
Perte de métal	3668	0	<b>3883</b>	587	247
Enfoncement	1641	5	12	<b>6815</b>	619
Délaminage	4448	0	5	1140	<b>11779</b>

TABLE 4.5 – Matrice de confusion résultat pour le SVM sur la base de tests avec augmentation artificielle du nombre de défauts

	Fausse alarme	Corrosion	Perte de métal	Enfoncement	Délaminage
Fausse alarme	<b>523</b>	1	5	5	31
Corrosion	35	<b>7</b>	2	4	1
Perte de métal	24	0	<b>50</b>	10	1
Enfoncement	8	1	6	<b>71</b>	6
Délaminage	26	0	1	9	<b>136</b>

TABLE 4.6 – Matrice de confusion résultat pour la forêt aléatoire sur la base de tests avec suppression de fausses alarmes

	Fausse alarme	Corrosion	Perte de métal	Enfoncement	Délamination
Fausse alarme	<b>51888</b>	0	13	26	68
Corrosion	41	<b>4908</b>	0	0	0
Perte de métal	6	0	<b>8579</b>	0	0
Enfoncement	0	0	0	<b>9292</b>	0
Délamination	14	0	0	0	<b>17358</b>

TABLE 4.7 – *Matrice de confusion résultat pour la forêt aléatoire sur la base de tests avec augmentation artificielle du nombre de défauts*

et pour les réseaux de neurones.

Les Figures 4.5 et 4.6 représentent respectivement les pourcentage de bonne classification (Nombre de défauts bien identifiés / Nombre de défauts présents dans la base) pour chaque type de défauts, avec rééquilibrage de la base de donnée par suppression de fausses alarmes et par augmentation artificielle du nombre de défauts.

Compte tenu de la difficulté de notre problématique, de tels résultats sont très satisfaisants et nous confortent dans le choix de nos caractéristiques : lors du déploiement industriel du logiciel complet, un bien plus grand nombre de défauts seront disponibles pour réaliser l'apprentissage, ce qui ne pourra être que bénéfique pour les performances résultantes, notamment sur les corrosions qui sont, comme l'on pouvait s'y attendre, les défauts les plus difficiles à classifier.

Du fait de la similarité des performances des forêts aléatoires et des réseaux de neurones si nous réduisent le nombre de fausses alarmes, et du fait que les performances, dans le cas du rééquilibrage par augmentation artificielle du nombre de défauts dans la base, sont très largement à l'avantage des forêts aléatoires, nous avons confirmé le choix de ces dernières pour le déploiement industriel.

## 4.5 Conclusion

Nous avons présenté dans cette étude une applications des méthodes de reconnaissances de formes sur une toute nouvelle problématique d'identification de défauts. Un grand nombre de caractéristiques morphologiques et *ad hoc* ont été exploitées, et notre base de tests à servi à la comparaison des classifieurs les plus communs utilisés de nos jours. Les résultats obtenus sont encourageants et nous confortent dans les choix réalisés, ouvrant la voie à un déploiement industriel de notre méthode à plus grande échelle.

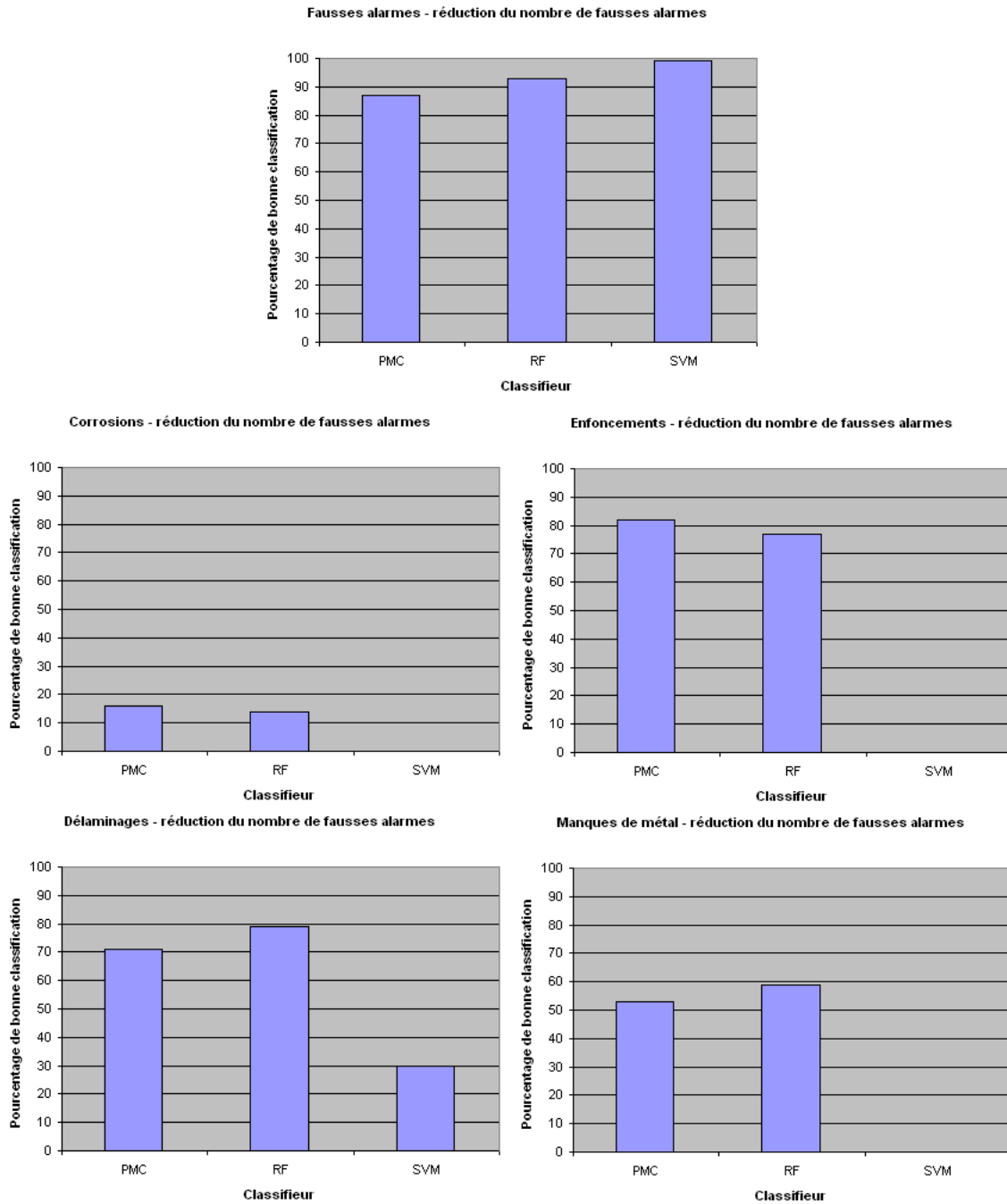


FIGURE 4.5 – Pourcentages de bonne détection pour chaque classes de défauts avec rééquilibrage de la base par suppression de fausses alarmes.

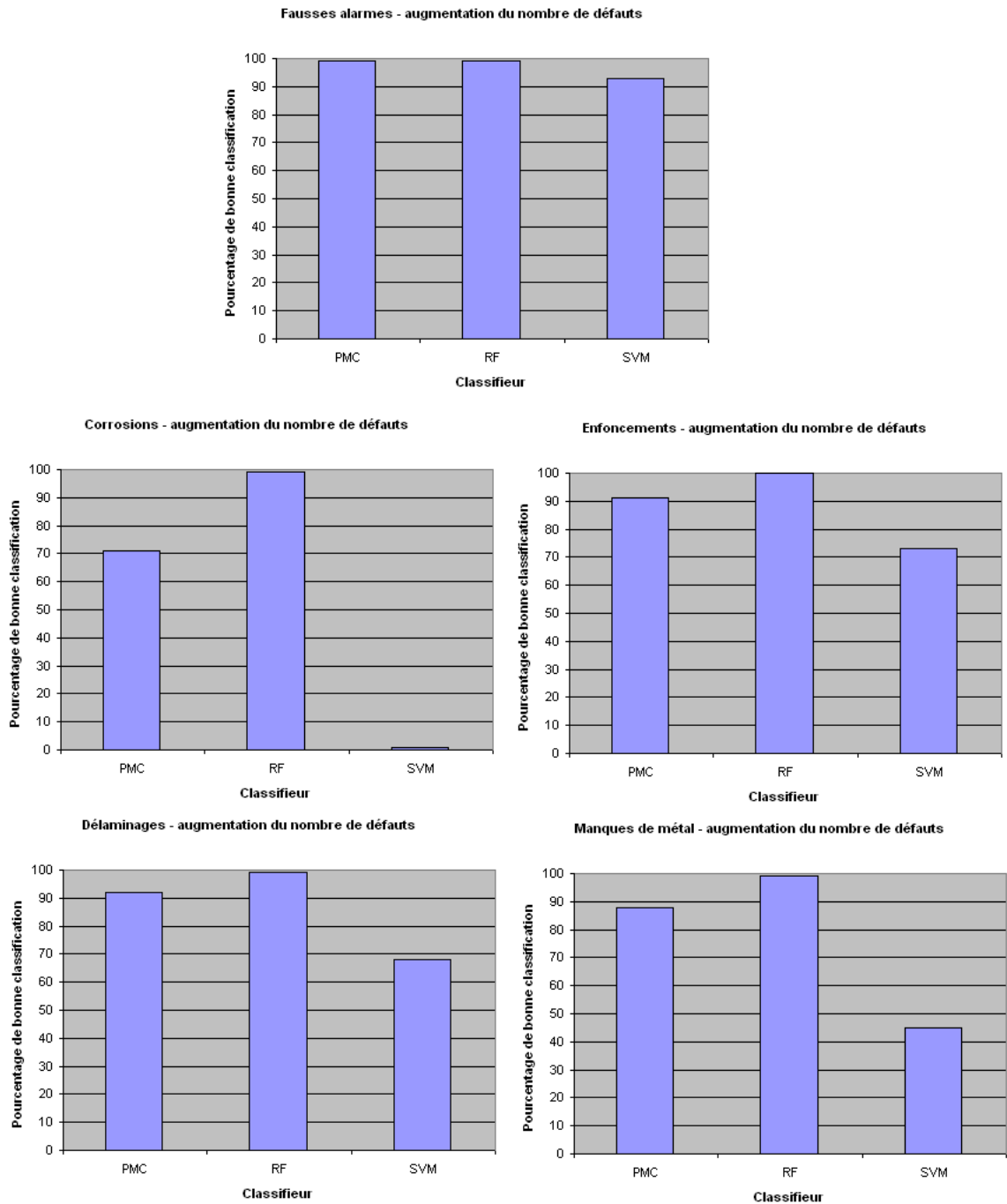


FIGURE 4.6 – Pourcentages de bonne détection pour chaque classes de défauts avec rééquilibrage de la base par augmentation artificielle du nombre de défauts.



# Conclusion générale

## Synthèse

Dans ce document nous avons présenté les diverses contributions apportées à l'automatisation de l'analyse de défauts sur les pipelines d'hydrocarbure par racleur instrumenté. Dans le second chapitre nous avons introduit une nouvelle méthode de détection des soudures d'aboutage se fondant sur l'utilisation de statistiques d'ordre élevé, à la fois performante (proches du 100%) et rapide (gain de 30% de temps d'exécution vis-à-vis de la méthode de référence utilisée jusque là).

Dans les chapitres suivants, nous avons présenté une chaîne complète de détection et d'identification de défauts au sein des images de coeur de tube (entre deux soudures d'aboutage). Une pré-segmentation des zones d'intérêt, utilisant une modélisation d'histogramme par algorithme EM adapté aux mélanges de gaussiennes, suivie d'une analyse statistique permet l'extraction des zones de danger, réduisant drastiquement (95% de compression) le nombre de pixels pris en compte tout en préservant une grande part (90%) des défauts effectivement présents et en particulier la quasi-totalité des plus importants en terme de dangerosité.

Enfin, nous avons présenté une étude des méthodes de reconnaissances de formes sur notre problématique, nouveau domaine d'application pour cette discipline. En résulte une chaîne complète d'analyse, pouvant se présenter comme un complément essentiel à l'analyse manuelle, à la fois performante et rapide d'exécution : aux vues des temps d'exécution MATLAB, nous estimons le temps d'exécution de la chaîne proposée, une fois codée en langage industriel, dans une fourchette de 30 à 40 heures pour une ligne de 30 kilomètres, un temps tout à fait raisonnable au vu de la quantité de données à traiter et en comparaison de l'actuel temps de traitement entièrement manuel (environ 15 jours/homme pour les mêmes 30 kilomètres).

## Déploiement industriel

Depuis le lancement du projet nous avons pu travailler main dans la main avec TRAPIL afin de répondre à leurs attentes, faisant évoluer les objectifs de la thèse au fur et à mesure des résultats obtenus. Depuis le développement et la mise en service de la méthode *READ-weld* en 2011, TRAPIL l'a intégré dans son processus d'analyse et l'exploite avec succès sur l'ensemble de ses nouveaux passages. Les résultats obtenus sur la détection et l'identification de défauts sont très encourageants et le développement en langage industriel (C++) est prévu courant 2014 pour une mise en service à la fin de cette année.

Ne pouvant se substituer à l'analyse manuelle, car le risque de rater un défaut est malgré tout présent, notre chaîne de détection va venir en complément de celle-ci, menant

à un pré-rapport automatisé. Celui-ci pourrait alors être transmis au client en quelques jours et lui permettre de commencer immédiatement sa campagne de fouilles, pendant que l'analyse manuelle complète est réalisée. Cela offrirait à TRAPIL une plus-value commerciale indéniable, offrant un premier livrable en l'espace d'une semaine, contre parfois un mois ou deux dans la situation actuelle.

## Perspectives

Le travail accompli est une première pierre sur cette nouvelle application qu'est le contrôle non destructif des pipelines d'hydrocarbure pour les méthodes de traitement de l'information ; de nombreux axes d'amélioration peuvent être suivis et il reste quelques points à approfondir.

En effet l'étude des soudures d'aboutage, ainsi que l'ensemble des défauts propres à cette zone particulière reste encore à développer, et nous disposons de nombreuses marges d'amélioration sur le travail accompli : l'étape de pré-détection est encore perfectible, il serait par exemple envisageable de combiner notre approche avec une étude des informations T1 afin de pouvoir détecter les enfoncements se trouvant dans une zone de bruit.

L'étape de classification pourrait grandement bénéficier d'une approche par cascade, en se focalisant tout d'abord sur une approche *One-Class* pour différencier dans un premier temps les défauts des fausses alarmes, puis d'entraîner un classifieur dédié à la différenciation des différentes classes de défauts.

Une autre possibilité pourrait être de renforcer la classification en lui appliquant une méthode de type boosting, notamment sur les classes de défauts les plus difficiles comme les corrosions.

Dans un autre domaine il serait possible d'embarquer certains traitements directement dans l'interface logicielle du racleur. La détection de soudures par exemple pourrait être réalisée au fur et à mesure de l'acquisition des données, notre approche très orientée signal ne nécessitant qu'un voisinage court des données. Cependant des difficultés se poseraient, notamment du fait que les données acquises par le racleur nécessitent un formatage pour être exploitées (étape de dépouillement), et que cela générerait un nouveau fichier de sortie à prendre en compte.

# Publications

## Prévues

### Revue internationale avec comité de lecture

- **Ultrasonics**, C. Fouquet, P. Duvaut et A. Histace. *Detection and Recognition of structural defects in oil pipelines using ultrasonic images*. Ultrasonics journal, prévu 2014

## En soumission

### Conférences internationales avec comité de lecture

- **ECNDT'14**, C. Fouquet, A. Histace et P. Duvaut. *Automated Classification of Defects Signature in Pipelines Using Ultra Sonic Thickness Images*. Conférence ECNDT 2014, Pragues, République Tchèque, Soumis

## Publiées

### Conférences internationales avec comité de lecture

- **NDT in progress'11**, C. Fouquet, P. Duvaut et D. Benboudjema. *Automatic and Circumferential Oriented Detection (ACOD) for Pipeline Intelligent Inspection*. Conférence NDT in progress 2011, Pragues, République Tchèque.
- **ICNDT'13**, C. Fouquet, A. Histace, P. Duvaut et L. Meziou. *Automated Detection Of Defects Signature in Pipelines Using Ultra Sonic Thickness Images*. Conférence ICNDT 2013, Portoroz, Slovénie.

### Rapports techniques internes

- **Comité de pilotage'12**, C. Fouquet, P. Duvaut. *Détection des soudures d'aboutage : méthodologie et résultats*. Comité interne, site de Poissy, Juin 2012.
- **Comité de direction'13**, C. Fouquet, P. Duvaut et A. Histace. *Détection de défauts automatique : méthodes explorées et planning*. Comité interne, Siege, Fevrier 2013.





# Bibliographie

- [1] M. Basseville and I. Nikiforov. Detection of abrupt changes : Theory and applications. *Englewood Cliffs : Prentice-Hall*, 1993.
- [2] L. Breiman. Random forests. *Machine learning*, 2001.
- [3] L. Breiman and J.H. Olshen. Classification and regression trees. *Chapman and Hall*, 1984.
- [4] C. Carson. Blobworld : Image segmentation using expectation maximization algorithm and its applications to image querying. *Pattern analysis and machine intelligence 24*, 2002.
- [5] A.A. Carvalho and J.M.A. Rebello. Mfl signals and artificial neural networks applied to detection and classification of pipe weld defects. *NDTandE International 39*, 2006.
- [6] D.A. Clausi. An analysis of co-occurrence texture statistics as a function of gray level quantization. *Can. J. Remote sensing 28*, 2002.
- [7] Pipeline Operator Forum Draft. *Specifications and requirements for intelligent pig inspection of pipelines*, 2009.
- [8] Y. Freund and R.E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Computational learning theory : Eurocolt*, 1995.
- [9] R.M. Haralick and K. Shanmugam. Textural features of image classification. *IEEE Transaction on systems, man and cybernetics 6*, 1973.
- [10] R. Huang and L.W. SchmerrJr. Characterization of the system functions of ultrasonic linear phased array inspection systems. *Ultrasonics 49*, 2008.
- [11] S. Iyer and S.K. Sinha. Ultrasonic signal processing methods for detection of defects in concret pipes. *Automation in Construction 22*, 2011.
- [12] M. Juneja and P.S. Sandhu. Performance evaluation of edge detection technique for images in spatial domain. *International Journal of Computer Theory and Engineering*, 1(5), dec 2009.
- [13] H.W. Kim and J.K. Lee. Circumferential phased array of shear-horizontal wave magnetostrictive patch transducers for pipe inspection. *Ultrasonics 53*, 2012.
- [14] N. Kumar. Fast parrallel expectation maximization for gaussian mixture models on gpu using cuda. *High performance computing and communication*, 2009.
- [15] M.J.S lowe and D.N. Alleyne. Defect detection in pipes using guided waves. *Ultrasonics 36*, 1998.
- [16] Kass M., Witkin A., and Terzopoulos D. Snakes : active contour models. *International Journal of Computer Vision*, 1, 1988.
- [17] S. Mallat. A wavelet tour of signal processing. *Academic Press*, 1999.

- [18] T.K. Moon. The expectation-maximization algorithm. *Signal processing magazine* 13, 1996.
- [19] S. Mori and C.Y. Suen. Historical review of ocr research and developpement. *Proceedings of the IEEE* 80, 1992.
- [20] H.B. Demuth M.T. Hagan. Neural network design. 1996.
- [21] S. Mukhopadhyay and G.P. Srivastava. Characterisation of metal loss defects from magnetic flux leakage signals with discrete wavelet transform. *NDTandE International* 33, 2000.
- [22] University of Waikato. [www.cs.waikato.ac.nz/ml/weka](http://www.cs.waikato.ac.nz/ml/weka).
- [23] N. Otsu. A threshold selection method for gray scale histogram. *IEEE Transaction on systems, man and cybernetics* 8, 1978.
- [24] M.A. Richards. Fundamentals of radar signal processing. *Mc-Graw-Hill Inc*, 2005.
- [25] R.S.C. Cobbold R.K. Warriner. Focused, phased-array plane piston and spherically-shaped concave piston transducers : Comparison for the same aperture and focal point. *Ultrasonics* 52, 2011.
- [26] M.S. Safizadeh and T. Azizzadeh. Corrosion detection of internal pipeline using ndt optical inspection system. *NDTandE International* 52, 2012.
- [27] B. Schölkopf and A. Smola. Advances in kernel methods - support vector learning. *MIT Press*, 1998.
- [28] H.I. Shafeek and E.S. Gadelmawla. Automatic inspection of gas pipeline welding defects using an expert vision system. *NDTandE International* 37, 2004.
- [29] L. Soh and C. Tsatsoulis. Texture analysis of sar sea ice imagery using gray level co-occurrence matrices. *IEEE Transaction on geoscience and remote sensing* 37, 1999.
- [30] H. Van Trees. Detection, estimation and linear modulation theory, part 1. *John Wiley*, 2001.
- [31] P. Viola and M. Jones. Fast and robust classification using asymmetric adaboost and a detector cascade. *Neural information processing systems* 14, 2001.
- [32] Y. Wang and Y. Sun. Detectin of line weld defects based on multiple thresholds and support vector machine. *NDTandE International* 41, 2008.
- [33] G. Winkler. Image analysis, random fields and markov chain monte carlo methods : A mathematical introduction. *Springer*, 2006.
- [34] S. Wold and K. Esbensen. Principal component analysis. *Chemometrics and intelligent laboratory systems* 2, 1987.
- [35] L. Xu and M.I. Jordan. On convergence properties of the em algorithm for gaussian mixtures. *Neural computation* 8, 1996.