



**HAL**  
open science

# Estimation de la qualité de bois ronds et d'arbres sur pied par Lidar terrestre

van Tho Nguyen

► **To cite this version:**

van Tho Nguyen. Estimation de la qualité de bois ronds et d'arbres sur pied par Lidar terrestre. Sylviculture, foresterie. AgroParisTech, 2018. Français. NNT : 2018AGPT0007 . tel-01904642

**HAL Id: tel-01904642**

**<https://hal.science/tel-01904642>**

Submitted on 25 Oct 2018

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

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

N° : 2018 AGPT 0007

## Doctorat ParisTech

### THÈSE

pour obtenir le grade de docteur délivré par

## L'Institut des Sciences et Industries du Vivant et de l'Environnement (AgroParisTech)

Spécialité : Biologie végétale et forestière

*présentée et soutenue publiquement par*

**Van-Tho NGUYEN**

le 10 octobre 2018

## Estimation de la qualité de bois ronds et d'arbres sur pied par Lidar terrestre

Directeur de thèse : **Francis COLIN**

Co-encadrement de la thèse : **Thierry CONSTANT**

**Jury** (présidé par M. Cédric Pradalier)

**M. Cédric PRADALIER**, Prof, Georgiatech - DREAM lab, Metz

Rapporteur

**M. Robert COLLET**, MdC, Arts et Métiers - ENSAM, LABOMAP EA 3633, Cluny

Rapporteur

**M. Paul McLEAN**, PhD, Forest Research, Royaume-Uni

Examinateur

**Mme Hind BRIL-EL HAOUZI**, MdC, CRAN, Université de Lorraine, Nancy

Examinatrice

**Mme Isabelle DEBLED-RENNESON**, Prof, Loria, Université de Lorraine, Nancy

Co-encadrante

**Mme Alexandra BAC**, MdC, LSIS, Polytech Marseille, Marseille

Examinatrice

**M. Francis COLIN**, DR, SILVA, INRA, Nancy

Directeur de thèse

**M. Thierry CONSTANT**, CR, SILVA, INRA, Nancy

Co-Directeur de thèse, invité

**M. Alexandre PIBOULE**, CR, Département RDI, ONF, Nancy

Co-encadrant, invité

**M. Bertrand KERAUTRET**, Prof, LIRIS, Université de Lyon 2, Lyon

Co-encadrant, invité

**AgroParisTech**

Université de Lorraine, AgroParisTech, Inra, Silva  
route d'Amance, 54280 Champenoux, France



# Estimation de la qualité de bois ronds et d'arbres sur pied par Lidar terrestre

## Assessment of roundwood and tree quality from terrestrial Lidar data



# Remerciements

À Thiéry Constant pour avoir passé beaucoup du temps avec moi pour résoudre tous les problèmes rencontrés pendant ma thèse qu'ils soient scientifiques ou de la vie quotidienne.

À Francis Colin pour m'avoir encadré, conseillé, encouragé et aussi pour son humour.

À Isabelle Debled-Rennesson et Bertrand Kerautret pour m'avoir accueilli au laboratoire Loria Nancy et pour m'avoir encadré et conseillé.

À Alexandre Piboule pour m'avoir initié à la plateforme Computree et pour des conseils sur les algorithmes.

À Alexandria Bac et Franck Hétroy-Wheeler pour les conseils en tant que membre du comité de thèse.

À Cédric Pradalier et Robert Collet pour avoir bien voulu accepter la tâche de rapporteur de mon mémoire et pour les conseils données en retour.

Aux personnes qui ont contribué à la phase d'acquisition des données : Florian Vast pour sa disponibilité et pour avoir scanné la plupart des arbres utilisés dans cette thèse, Emmanuel Cornu pour m'avoir aidé à scanner des billons de merisier.

À Nathalie Morel, Elodie Taillefumier, Corinne Martin et Corinne Fiers pour le soutien administratif.

À tous ceux avec qui j'ai eu la chance de travailler ou de discuter : les chercheurs (Gérard Neuve, Julien Ruelle, Frédéric Mothe, Jana Dlouha, Yves Ehrhart, Fleur Longuetaud, Daniel Rittié, Ignacio Berbeito, ...) les (ex-)doctorants et post docs (Anjy, Citra, Jean-Léon, Lara, Félix, Vivien, Jialin, Estelle, Nicolas, Masoumeh, Raphaël, Mathieu, Laura, Gonzalo, ...) et les ingénieurs et techniciens (Phillipe Jaquin, Alain Mercanti, Etienne, Maryline Harroué, Etienne Farré, Charline Freyburger, Pierre Gelhaye, Adeline Motz, ...).

Au Labex Arbre et à la région Lorraine, qui m'ont permis, grâce à une allocation de recherches et diverses aides financières, de me consacrer sereinement à l'élaboration de ma thèse.

Et à ma femme, Thi pour son soutien très précieux de tous les instants.

# Table des matières

<b>Remerciements</b>	<b>iii</b>
<b>Liste des publications et communications</b>	<b>x</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Contexte . . . . .	1
1.2 La technologie LiDAR . . . . .	3
1.2.1 Le LiDAR terrestre . . . . .	3
1.2.2 Le LiDAR terrestre dans le domaine forestier . . . . .	4
1.3 Estimation de la qualité des bois en forêt . . . . .	6
1.3.1 Méthode conventionnelle pour évaluer la qualité du bois . . . . .	6
1.3.2 Estimation automatique de la qualité du bois en utilisant le LiDAR terrestre . . . . .	6
1.4 Défauts externes du bois à la surface d'un tronc . . . . .	10
1.4.1 Origine de défauts liés à la ramification . . . . .	10
1.4.2 Branche séquentielle . . . . .	11
1.4.3 Cicatrice de branche . . . . .	11
1.4.4 Amas de bourgeons . . . . .	12
1.4.5 Picot . . . . .	12
1.4.6 Pousse épïcormique . . . . .	13

1.4.7	Broussin . . . . .	14
1.4.8	Sphéroblaste . . . . .	15
1.5	Objectifs de la thèse . . . . .	16
<b>2</b>	<b>Détection des défauts à la surface du tronc</b>	<b>19</b>
2.1	Contexte et objectifs . . . . .	19
2.2	Méthodologie . . . . .	20
2.2.1	Calcul de la ligne centrale . . . . .	20
2.2.2	Calcul de la distance de référence à la ligne centrale . . . . .	22
2.2.3	Seuillage automatique . . . . .	22
2.2.4	Segmentation des branches et des défauts présentant un axe long . . . . .	23
2.3	Résultats principaux . . . . .	23
2.4	Conclusion . . . . .	24
2.5	Article 1 : Segmentation of defects on log surface from terrestrial Lidar data . . . . .	25
2.6	Article 2 : Algorithms and Implementation for Segmenting Tree Log Surface Defects	32
2.7	Article 3 : Branches and Defects Segmentation on the surface of tree trunk from partial or full terrestrial Lidar data. (Projet d'article, en cours) . . . . .	50
<b>3</b>	<b>Identification des défauts</b>	<b>68</b>
3.1	Contexte et objectifs . . . . .	68
3.2	Méthodologie . . . . .	69
3.3	Résultats principaux . . . . .	70
3.4	Conclusions et perspectives . . . . .	71
3.5	Article 4 : A machine-learning approach for classifying defects on tree trunk using terrestrial LiDAR (Projet d'article) . . . . .	73
<b>4</b>	<b>Caractérisation des défauts externes : vers une évaluation automatique de la qualité de</b>	

<b>l'arbre sur pied</b>	<b>95</b>
4.1 Contexte et objectifs . . . . .	95
4.2 Méthodologie . . . . .	96
4.3 Résultats principaux . . . . .	98
4.4 Intégration à la plate-forme Computree . . . . .	99
4.5 Conclusion . . . . .	100
4.6 Article 5 : An innovative and automated method for characterizing wood defects on trunk surface using high density 3D terrestrial LiDAR data . . . . .	102
<b>5 Discussion et perspectives</b>	<b>128</b>
5.1 Synthèse de la méthode et des résultats principaux . . . . .	128
5.2 Détection des défauts externes à partir des données TLS . . . . .	130
5.2.1 Résolution spatiale des données TLS et dimension des défauts détectables . . . . .	130
5.2.2 Jeux de données et temps de traitement . . . . .	131
5.2.3 Occlusion . . . . .	131
5.3 Techniques d'identification de défauts : comment améliorer? . . . . .	132
5.4 Influence de l'espèce en dehors des types de défauts . . . . .	133
5.5 Prise en compte des défauts non liés à la ramification . . . . .	133
5.6 Une analyse plus fine des défauts traités . . . . .	134
5.6.1 Classification des branches séquentielles et des branches épicorniques . . . . .	134
5.6.2 Classification des nœuds sains et des nœuds pourris . . . . .	135
5.7 Applications potentielles . . . . .	136
5.7.1 Un outil pour classer des grumes ou des arbres sur pied . . . . .	136
5.7.2 Adaptation pour les bois ronds . . . . .	136
5.8 Mise en relation avec les données de tomographie à rayons X . . . . .	137



**Conclusion**

**139**

**Bibliographie**

**142**

# Table des figures

1.1	Exemple d'utilisation de TLS dans le domaine forestier. . . . .	4
1.2	Illustration de la mesure de défauts sur un tronc par grimpage. . . . .	7
1.3	Illustration d'une branche séquentielle et des noeuds dans le bois dus aux branches séquentielles. . . . .	11
1.4	Illustrations de cicatrices de branche sur le tronc d'espèces différentes. . . . .	12
1.5	Illustrations d'amas sur les troncs de chêne. . . . .	12
1.6	Illustration d'un picot du chêne. . . . .	13
1.7	Illustrations de pousses épicorniques. . . . .	13
1.8	Illustrations de broussins sur les troncs de chênes : (a) un broussin composé d'un gourmand et des bourgeons, (b) un broussin composé de poils et de bourgeons et (c) un broussin composé uniquement de bourgeons. . . . .	14
1.9	Illustration des traces d'un broussin : "patte de chat". . . . .	14
1.10	Illustration d'un sphéroblaste. . . . .	15
1.11	Schéma de l'approche en 3 étapes. . . . .	17
1.12	Description du matériel végétal utilisé dans les expérimentations à chaque étape. . . . .	17
2.1	Exemple de la difficulté d'ajustement d'un cercle à la section transversale d'un tronc avec une forme irrégulière. . . . .	21
2.2	Schéma pour la segmentation des défauts . . . . .	21

2.3	Un patch longitudinal du point $P$ (a) et l'ajustement d'une ligne au patch pour calculer la distance de référence à la ligne centrale ( $\hat{r}$ ) (b). L'axe $z$ représente la position longitudinale dans une orientation globale du billon selon sa longueur. . . .	22
3.1	Schéma du principe d'une méthode d'apprentissage automatique supervisée . . . .	69
3.2	Visualisation des résultats de l'identification de types des défauts sur les maillages du Chêne 411 (a) et du Hêtre 103 (b). ■ représente un défaut de type branche, ■ représente une cicatrice de branche, ■ représente un broussin et ■ représente un petit défaut. . . . .	71
4.1	Description du calcul de la position du défaut $D$ par les coordonnées de son centre $P$ , les dimensions de $D$ sont déterminées par sa boîte englobante (a). Une cicatrice de branche se caractérise par une cicatrice elliptique et une moustache (b). . . . .	96
4.2	Illustration du calcul des dimensions de la cicatrice de la branche en utilisant le profil des extensions verticales du défaut le long de l'arc périphérique. . . . .	97
4.3	Diagrammes en violon des différences entre l'estimation et l'observation des dimensions du défaut pour différents types de défauts. Pour chaque diagrammes, les lignes horizontales indiquent les quartiles et la médiane, cette dernière étant accompagnée par sa valeur. . . . .	98
4.4	Capture d'écran du plugin dans le plateforme Computree . . . . .	100
5.1	Une maladie du T sur un hêtre (a) et une fente sur un chêne détectées par l'algorithme proposée. . . . .	134
5.2	Différence en forme de la cicatrice d'une branche séquentielle (a) qui a une forme de "moustache de Chinois" et d'une branche épïcormique (b) qui a une forme ronde. . . . .	135

# Liste des publications et communications

## Publications publiées

- Nguyen VT, Kerautret B, Debled-Rennesson I, Colin F, Piboule A, Constant T (2016c) Segmentation of defects on log surface from terrestrial lidar data. In : Pattern Recognition (ICPR), 2016 23rd International Conference on, IEEE, pp 3168–3173
- Nguyen VT, Kerautret B, Debled-Rennesson I, Colin F, Piboule A, Constant T (2016a) Algorithms and implementation for segmenting tree log surface defects. In : International Workshop on Reproducible Research in Pattern Recognition, Springer, pp 150–166

## Publications non-publiées

- Nguyen VT, Kerautret B, Debled-Rennesson I, Colin F, Piboule A, Constant T (2018c) Branches and defects segmentation on the surface of tree trunk from partial or full terrestrial lidar data, (en projet)
- Nguyen VT, Constant T, Kerautret B, Debled-Rennesson I, Colin F (2018b) A machine-learning approach for classifying defects on tree trunk using terrestrial LiDAR, (à soumettre)
- Nguyen VT, Constant T, Colin F (2018a) An innovative and automated method for characterizing wood defects on trunk surface using high density 3d terrestrial lidar data, (soumis)

## Communications orales

- V-T. Nguyen, T. Constant, B. Kerautret, I. Debled-Rennesson, B. Garnier, A. Piboule, F. Colin. Oak epicormics : feasibility of detection by terrestrial Lidar. EPIC-IUFRO, 15-16 Septembre 2017, Nancy, France
- V-T. Nguyen, B. Kerautret, I. Debled-Rennesson, F. Colin, A. Piboule, T. Constant. De la description 3D des bois ronds à la caractérisation des défauts internes : nouveaux savoir-faire et perspectives. Caqsis, 30 Mars 2017, Bordeaux, France
- V-T. Nguyen, B. Kerautret, I. Debled-Rennesson, F. Colin, A. Piboule, T. Constant. Détection

de défauts à la surface du tronc d'arbres à partir de données T-Lidar. Atelier Lidar, 23-24 novembre 2016, Avignon, France

## Posters

- V-T. Nguyen, B. Kerautret, I. Debled-Rennesson, F. Colin, A. Piboule, T. Constant. Segmentation of defects on log surface from terrestrial Lidar data. Pattern Recognition (ICPR) "23rd International Conference on Pattern Recognition", 4-8 décembre 2016, Cancun, Mexique
- V-T. Nguyen, B. Kerautret, I. Debled-Rennesson, F. Colin, A. Piboule, T. Constant. Algorithms and Implementation for Segmenting Tree Log Surface Defects. 1st Workshop on Reproducible Research in Pattern Recognition, 4 décembre 2016, Cancun, Mexique
- V-T Nguyen, T. Constant, D. Elhareth, A. Piboule, F. Colin. Analyse de la surface de bois rond pour la détection de défauts internes. GDR-Bois, 4-6 novembre 2015, Clermont-Ferrand, France
- V-T Nguyen, T. Constant, D. Elhareth, A. Piboule, F. Colin. Analysis of roundwood surface for detection of inner defects. Silvilaser, 28-30 septembre 2015, La Grande Motte, France



# Chapitre 1

## Introduction

### 1.1 Contexte

Le bois est une matière première d'origine biologique et non homogène. Sa qualité est très variable. Elle résulte des interactions entre des facteurs génétiques et environnementaux modifiés par la sylviculture. Au delà de ces critères, la qualité d'un arbre s'évalue à partir de ses dimensions et de la présence, de la distribution, du nombre et des dimensions de ses défauts. En général, elle décroît avec le nombre et la dimension des défauts. Cependant, une petite trace noire de quelques mm due à un picot a un impact plus important qu'un plus grand nœud dû à un gourmand. Enfin, pour les essences dont l'aubier est mis de côté, si le défaut se trouve dans l'aubier, l'impact est moins important (voir plus de détails à la section 1.4). Par contre, si sa présence est dans le duramen, l'impact sur la qualité est plus grave, car c'est la partie utilisée par la filière.

L'utilisation du bois comme matériau dépend de sa qualité. Le prix du bois dépend donc fortement de sa qualité. Chez le chêne, le prix varie de 4,5 à 6,5 fois entre un billon de classe A et un billon de classe B du classement comprend 4 classes (Jourez et al., 2010). Le classement de la qualité des arbres ou des bois ronds est donc une tâche importante dans l'inventaire forestier ou avant la récolte (Fonseca, 2005) pour maximiser la valeur de la vente. Elle est aussi importante pour optimiser le processus du sciage des bois ronds dans les scieries (Chang, 1992).

Afin de classer les arbres ou les bois ronds, il faut trouver les défauts, puis les caractériser. Différentes méthodes ont été proposées. On peut classer ces méthodes en deux catégories. La première catégorie classe les bois par les défauts internes tandis que la deuxième utilise les défauts externes.

Plusieurs techniques sont développées pour la détection de défauts internes telles que la MRI (Magnetic Resonance Imaging), l'échographie (Schafer, 2000) et la tomographie à rayons X. La dernière est la plus souvent utilisée, car elle permet de détecter les nœuds ou d'autres traces

de ramification avec une très grande précision jusqu'à 1 mm (Colin et al., 2010b). Jusqu'à 95 % de défauts ont pu être détectés grâce à cette technologie (Li et al., 1996). Des applications ont été développées pour permettre une détection automatique des nœuds dans les bois (Krähenbühl et al., 2012, 2016) à partir de données 3D venant de la tomographie à rayons X. Des chaînes de traitement utilisant cette technologie sont disponibles dans les scieries. Cependant, cette technologie exige que l'arbre soit abattu et coupé en billons ce qui limite donc son utilisation. Un autre désavantage de la tomographie à rayons X est le coût d'équipement souvent très élevé. Seules les plus grandes scieries ont les moyens pour investir dans une telle technologie (Thomas et al., 2007).

La technologie LiDAR (Light Detection And Ranging) en particulier le LiDAR terrestre (TLS) apparaît comme une approche plus abordable que la tomographie à rayons X pour une utilisation dans le domaine forestier. Cette technologie permet de numériser des scènes forestières ou des arbres individuels avec beaucoup de détails et une grande précision. De plus en plus de recherches ont été réalisées pendant la dernière décennie pour tirer profit de cette technologie dans le domaine de l'inventaire forestier. Récemment, des recherches ont montré qu'il y a une forte corrélation entre les défauts internes et externes (Thomas, 2009; Stängle et al., 2013; Racko, 2013; Pyörälä et al., 2017). En comparant le résultat de la mesure sur les données TLS et sur les données de tomographie à rayons X, toutes ces études ont obtenu des coefficients de détermination supérieurs à 0.6. Cette forte corrélation suggère que les défauts externes peuvent être utilisés pour estimer la qualité des bois ronds ou des arbres sur pied. Des recherches sur la détection de défauts externes ont été menées (Schütt et al., 2004; Thomas et al., 2007; Kretschmer et al., 2013; Stängle et al., 2013) pour but d'estimer la qualité du bois rond. Cependant, ces recherches se sont concentrées sur de larges défauts (Thomas et al., 2007) ou sont semi-automatiques (Kretschmer et al., 2013; Stängle et al., 2013; Schütt et al., 2004). Il manque par conséquent des études sur la détection, la classification et la caractérisation automatique des défauts. De plus, les travaux existants étaient sur une ou deux espèces particulières et concernaient principalement les cicatrices de branche. Bien que la cicatrice de branche soit un des défauts les plus fréquents, les autres défauts tels que les broussins, amas et picots sont aussi importants si on veut construire un système de décision pour remplacer une procédure conventionnelle de classement des bois ronds (AFNOR, 2012a,b, 1999a,b,c) ou de l'arbre sur pied (Jourez et al., 2010).

Cette thèse vise à développer une méthode automatique pour détecter, classer et caractériser les défauts externes du tronc vers une estimation automatique de la qualité de l'arbre sur pied ou du bois rond.



## 1.2 La technologie LiDAR

La technologie LiDAR (Light Detection and Ranging) est une technologie de numérisation fondée sur le principe de la télémétrie. Il s'agit d'un système de détection actif qui émet des faisceaux laser intenses et focalisés vers une cible et capture la réflexion par la suite. Deux méthodes sont utilisées pour mesurer la distance du capteur à l'objet cible : le décalage de phase et le temps de vol. La première méthode consiste à mesurer le décalage de phase entre le signal émis et le signal réfléchi. Une seule impulsion de retour est enregistrée pour chaque direction d'émission. Les scanners qui tirent profit du décalage de phase permettent de numériser un nombre très important de points en un temps court d'acquisition, mais sont assez limités en portée (quelques centaines de mètres au maximum). La seconde méthode consiste à mesurer le temps nécessaire pour que la lumière émise atteigne l'objet cible et revienne au capteur. Cette méthode permet des mesures sur des objets à une distance plus grande mais avec une vitesse d'acquisition plus faible par rapport à la méthode du décalage de phase.

Le scanner LiDAR produit un nuage de points 3D de la scène mesurée. Les coordonnées de chaque point  $(x, y, z)$  sont calculées à partir de :

- 1) la distance
- 2) la direction visée
- 3) la location absolue du capteur.

En plus de l'information sur les coordonnées 3D de l'objet, le LiDAR fournit souvent une information sur l'intensité du retour associée à chaque point. L'intensité d'un point dans le nuage de points dépend de trois facteurs : (1) l'angle d'incidence entre le faisceau laser et la surface de l'objet, (2) les propriétés optiques de l'objet, et (3) la distance entre l'objet et le capteur. L'intensité peut être utilisée comme un complément pour la détection de défauts (Schütt et al., 2004).

### 1.2.1 Le LiDAR terrestre

Un scanner LiDAR terrestre (TLS) est un système de numérisation le plus souvent hémisphérique. Il contient un miroir rotatif pour un scan circulaire selon un angle zénithal ( $\theta$ ) combiné à une rotation azimutale ( $\phi$ ). Le mouvement dans ces deux directions permet une vue hémisphérique complète à l'exception de la zone masquée par la base de l'instrument formant une région occluse d'environ 20 degrés en dessous. Les scanners TLS sont des systèmes de numérisation à courte portée. Ils arrivent à numériser des objets à des distances allant de 1 à quelques kilomètres, selon le modèle et la technique pour mesurer la distance. Ils ont une précision de l'ordre de quelques millimètres. Les scanners TLS peuvent s'installer au sol et être montés sur une plate-forme fixe comme un trépied (La Figure 1.1 (b)).



(a) Illustration de scanner Faro Focus 3D X130 avec la rotation zénithale ( $\theta$ ) et la rotation azimutale ( $\phi$ )



(b) Scène d'un scan TLS

**Figure 1.1** Exemple d'utilisation de TLS dans le domaine forestier.

## 1.2.2 Le LiDAR terrestre dans le domaine forestier

Dans le domaine forestier, en numérisant une scène forestière en 3D, le TLS peut fournir des informations sur une parcelle ou un arbre individuel. Dans ces dernières décennies, des applications ont été développées pour exploiter la richesse des données récupérées du TLS. Les premières études ont exploité le potentiel de remplacement de méthodes conventionnelles de mesure des paramètres de l'inventaire forestier telles que la hauteur des arbres, le diamètre à hauteur de poitrine (DHP - diamètre du tronc mesuré à 1,3 m au-dessus du niveau du sol), la densité à l'hectare du nombre de tiges et le volume de bois ou la biomasse. Au niveau de la parcelle, la tâche principale est la détection des arbres ou des tiges (Olofsson et al., 2014; Litkey et al., 2008; Liang et al., 2008; Bienert et al., 2007; Popescu et al., 2002; Hopkinson et al., 2004; Henning et Radtke, 2006a; Maas et al., 2008; Heinzl et Huber, 2016; Simonse et al., 2003). À partir des arbres ou des tiges détectés, des informations utiles pour l'inventaire peuvent être déduites. Une reconstruction d'arbres ou de tiges est ensuite faite pour mesurer le DHP (Olofsson et al., 2014; Litkey et al., 2008; Liang et al., 2008; Bienert et al., 2007; Simonse et al., 2003), la hauteur des arbres (Popescu et al., 2002) ou à la fois le DBH et la hauteur (Hopkinson et al., 2004; Henning et Radtke, 2006a; Maas et al., 2008; Moskal et Zheng, 2011; Calders et al., 2015). D'autres paramètres de la

parcelle peuvent être calculés à partir du TLS tels que le volume de bois (Kwak et al., 2014; García et al., 2011), la densité des tiges (Watt et Donoghue, 2005; Brolly et Király, 2009) et les structures de la forêt telles que la mesure observée ou modélisée de la canopée (Zande et al., 2008; Seidel et al., 2015; Cifuentes et al., 2017; Henning et Radtke, 2006b; Danson et al., 2006). Le TLS est aussi utilisé pour caractériser les changements de la forêt entre deux dates (Liang et al., 2012) ou le suivi (Vastaranta et al., 2012).

Afin de tirer avantage du détail et de la précision de données TLS, de nombreuses recherches ont été réalisées pour mesurer des caractéristiques d'un arbre individuel. Les objectifs de ces études sont d'abord la reconstruction (Pfeifer et al., 2004; Kelbe et al., 2013, 2015; Wang et al., 2016) ou la modélisation (Raumonen et al., 2013; Côté et al., 2011) de l'arbre. Des caractéristiques de tige sont extraites telles que le DHP (You et al., 2016), le DHP et la hauteur de l'arbre (Olofsson et al., 2014). La mesure du volume du bois ou de la biomasse de l'arbre individuel est souvent analysée (Hauglin et al., 2013; Dassot et al., 2012; Yu et al., 2013; Kankare et al., 2013; Holopainen et al., 2011; Bienert et al., 2014; Lefsky et McHale, 2008). Le LAI (leaf area index) peut aussi être calculé à partir de données TLS (Moorthy et al., 2008). Dans le cadre des mesures d'inventaire, des recherches ont été consacrées à la comparaison de la mesure utilisant le TLS avec les mesures par des méthodes conventionnelles qui ont montré l'efficacité de la méthode TLS (Mengesha et al., 2015; Hopkinson et al., 2004). Par ailleurs, grâce au TLS, des propriétés de qualité de la tige peuvent être mesurées telles que la courbure générale (Liang et al., 2014a), le défilement et les courbures locales (Van Leeuwen et al., 2011; Kankare et al., 2014). Klemmt et al. (2010) ont proposé une méthode pour mesurer la hauteur des verticilles ainsi que le nombre de branches par verticille du Pin maritime.

Le grande précision des données de TLS ouvre de nouvelles possibilités de décrire la texture ou la rugosité de l'écorce pour de nouvelles applications de TLS. Othmani et al. (2013) ont utilisé une technique de l'analyse de texture de l'écorce pour reconnaître l'espèce de l'arbre. L'analyse de la rugosité de l'écorce permet de détecter les défauts externes du tronc (Schütt et al., 2004; Thomas et al., 2007, 2010; Kretschmer et al., 2013; Stängle et al., 2013). Une revue plus détaillée de cette dernière utilisation du TLS est présentée dans la section suivante.

## 1.3 Estimation de la qualité des bois en forêt

### 1.3.1 Méthode conventionnelle pour évaluer la qualité du bois

L'évaluation de la qualité du bois est une tâche importante dans l'inventaire forestier et avant la récolte des arbres pour maximiser la valeur marchande. Extérieurement, la qualité du bois d'un tronc ne peut pas être évaluée seulement par ses diamètres, sa courbure et son fil tors. La qualité est également affectée par les défauts selon leurs emplacements, types, nombre et dimensions. Par exemple, les défauts epicormiques (gourmands, broussins et picots) peuvent dégrader la qualité des chênes au moins d'une classe (Meadows et Burkhardt, 2001). Les bois ronds ou des billons de l'arbre sur pied sont classés selon leur utilisation. Le classement qualitatif des bois ronds feuillus selon les normes européennes (AFNOR, 2012a,b) comporte 4 classes de qualité. La classe A est une classe de qualité exceptionnelle. Les bois ronds de la classe A sont destinés au tranchage ou au déroulage pour produire des placages. Ils sont également destinés à la confection des merrains pour les chênes. Cette classe de qualité exige qu'il n'y ait aucun défaut sur un billon de hêtre (AFNOR, 2012a; Jourez et al., 2010) ou un seul petit défaut (un nœud sain avec un diamètre de moins de 15 mm ou un picot) sur un billon de chêne (AFNOR, 2012a; Jourez et al., 2010). Les bois de la classe B sont de 1ère et 2ème qualités et peuvent être utilisés en ébénisterie et menuiserie (cuisines) et pour la fabrication des merrains. L'exigence de défauts est moins importante que pour la classe A. Cependant, cette classe n'autorise pas de gros défauts, par exemple, les défauts doivent avoir un diamètre de moins de 60 mm et la somme des diamètres des défauts doit être inférieure à 100 mm (AFNOR, 2012a). La classe C fournit les avivés et des bois de charpente. Les nœuds sains sont admis mais le diamètre des nœuds pourris ne doit pas dépasser 50 mm. La classe D est une classe de qualité industrielle (traverses, palettes) et pour le bois d'énergie. Il n'y a pas de contrainte de défauts pour cette classe. L'évaluation de la qualité d'arbre pour appliquer cette méthode nécessite une grande expertise de la part d'un forestier spécialisé dans le classement. Une autre façon, plus utilisée par les services de recherche forestière, est de grimper sur le tronc pour repérer et mesurer chaque défaut (Figure 1.2).

### 1.3.2 Estimation automatique de la qualité du bois en utilisant le LiDAR terrestre

L'avantage du TLS est sa capacité à mesurer l'arbre sur pied avec beaucoup de détails et une grande précision, ce qui offre de nouvelles possibilités pour estimer la qualité du bois d'arbre sur pied ou de bois rond. Grâce à ces propriétés, le TLS permet une analyse détaillée de la rugosité ou de la texture de la surface du billon ou du tronc et permet de détecter les défauts notamment ceux liés à la ramification. L'apparence de ces défauts à la surface du tronc dépend de plusieurs facteurs tels que l'âge, l'espèce de l'arbre et le type de défaut. En général, ce sont des zones



**Figure 1.2** Illustration de la mesure de défauts sur un tronc par grimpage.

saillantes (Carpenter et Jones, 1989) par rapport à la surface. Autrement dit, pour détecter ces défauts externes, il est nécessaire de détecter de changements locaux de relief de la surface du tronc. Les travaux existant sur la reconstruction de la surface du tronc sont basés sur l'ajustement de forme primitive en presumant qu'une coupe transversale du tronc est parfaitement circulaire. À partir de cette hypothèse, il y a deux approches principales pour la reconstruction du tronc. La première se base sur un ajustement de cylindres (Pfeifer et al., 2004). La deuxième se base sur un ajustement de cercles (Bienert et al., 2007; Thomas et Mili, 2007). Pfeifer et al. (2004) ont proposé un ensemble d'algorithmes pour l'ajustement et le suivi des cylindres de la tige et de branches à partir de données LiDAR terrestre. Bienert et al. (2007) ont utilisé la méthode des moindres carrés pour l'ajustement des cercles, ce qui permet d'estimer le diamètre du tronc à différentes hauteurs. Thomas et Mili (2007) ont développé une nouvelle méthode "GM Estimator" pour ajuster des cercles, qui est plus robuste que la précédente vis-à-vis du bruit du nuage de points.

En se basant sur la reconstruction de la surface du tronc ou du billon évoquée ci-dessus, plusieurs recherches utilisant la détection des défauts pour estimer de la qualité du bois ou de l'arbre ont été menées (Schütt et al., 2004; Thomas et al., 2007, 2010; Thomas et Thomas, 2013; Kretschmer et al., 2013; Stängle et al., 2013).

Schütt et al. (2004) sont les premiers à avoir montré que les défauts du bois peuvent être localisés en analysant l'image 2D correspondant au nuage de points 3D du LiDAR terrestre. La procédure pour passer du nuage de points 3D en images 2D consiste d'abord à construire des cylindres successifs correspondant à la surface du tronc. Ensuite, le nuage de points est converti du système de coordonnées cartésiennes vers un système de coordonnées cylindriques selon les cylindres ajustés. Deux images 2D sont ensuite construites en déroulant le nuage de points associé à chaque

cylindre. La première est une image de profondeur (range image en anglais). La deuxième image est une image d'intensité. Un réseau de neurones est entraîné par des caractéristiques extraites de ces images. Une application a été développée qui permet à l'utilisateur de corriger les résultats de la reconnaissance automatique. Cependant, les auteurs n'ont pas décrit en détails comment extraire les caractéristiques, ni les modalités ni le résultat de leur classification.

Thomas et al. (2007, 2010) ont utilisé cependant une autre approche. D'abord une image 2D est créée à partir des ajustements de cercles sur les sections transversales du billon par un estimateur robuste (Thomas et Mili, 2007). La valeur d'un pixel de l'image 2D est le rayon du cercle correspondant. Les défauts de type cicatrice de branche sont détectés par une analyse de contours issus du relief dans des images 2D prises à six niveaux de hauteur. Le résultat du test sur les billons de chêne rouge et de tulipier (yellow poplar) a montré qu'on peut détecter de grands défauts à la surface à partir de ces images (Thomas et al., 2007). En effet, 63 défauts sur le total de 68 ont été détectés avec 10 fausses détections. La dimension de ces défauts était supérieure à 12,7 cm en diamètre et avec un relief supérieur à 2,5 cm. Thomas et al. (2010) ont ajouté une interface graphique au travail précédent. Ils ont aussi amélioré leur travail précédent en traitant le nuage de points en parallèle (Thomas et Thomas, 2013). Le tronc d'arbre est coupé en tronçons, puis ces derniers sont traités en parallèle. La parallélisation permet de réduire le temps de calcul. Le temps moyen est descendu à 1,009 secondes par rapport au 49,973 de celui de l'algorithme non parallélisé. Ils ont montré que le taux de détection était meilleur mais que les faux positifs (les défauts détectés ne correspondant à aucun défaut) étaient plus nombreux.

Kretschmer et al. (2013) ont modélisé l'écorce du tronc d'arbre en 3D pour faciliter la mesure des branches et les cicatrices de branches sur l'écorce d'arbres sur pied. La surface de référence du tronc est reconstruite par l'ajustement de cylindres consécutifs (Pfeifer et al., 2004). Ensuite, le nuage de points est converti en système de coordonnées cylindriques. Une carte de distance est construite dont la valeur est la différence entre les points 3D et la surface de référence. Dans cette carte, la couleur des pixels est un dégradé lié à la différence. Grâce à cette carte, les régions de défauts correspondant au fort relief sont colorées en rouge et sont mise en évidence pour un utilisateur qui peut ensuite mesurer leur dimension. Leurs résultats sont comparables à ceux obtenus par la mesure manuelle sur arbre abattu. En effet, 17 sur 18 cicatrices de branche ont été détectées sur 3 hêtres avec 58 % des défauts qui ont une erreur de mesure inférieure à 1 cm.

Stängle et al. (2013) ont utilisé une approche similaire à celle de Kretschmer et al. (2013) pour estimer le volume de bois sans nœud du hêtre. Ils ont aussi démontré que certains attributs de l'écorce externe peuvent être reliés à la qualité interne des billons en couplant l'analyse de leur forme avec des données provenant de tomographie à rayons X. Le résultat a montré qu'il y a une corrélation moyenne entre l'évaluation de la qualité réalisée par LiDAR terrestre et celle réalisée par la tomographie à rayons X. Sur 85 billons, 62,5 % ont eu la même classe de qualité selon les deux méthodes.

Hors du domaine forestier, plusieurs méthodes et techniques ont été développées afin de détecter les défauts d'une surface des murs (Laefer et al., 2014) et des avions (Jovančević et al., 2017). Bon nombre de méthodes se basant sur les techniques de traitement d'images et d'apprentissage automatique ont aussi été proposées pour la détection de défauts de surface. Suvdaa et al. (2012) ont utilisé la méthode SVM (Cortes et Vapnik, 1995) d'apprentissage supervisée, la méthode SIFT basé sur des descripteurs invariants (Lowe, 2004) afin de détecter les défauts de la surface des aciers. Sinha et Fieguth (2006) ont utilisé le réseau de neurones flous pour détecter et classifier les défauts de tuyaux. Les défauts à la surface de tuyaux sont très variables. Les descripteurs (pour l'extraction de caractéristiques) se basent sur 5 critères : intensité, texture, taille, forme et organisation (Azimi-Sadjadi et al., 1996). López-García et al. (2010) ont proposé une méthode pour détecter les défauts visuels de citrons. Cette approche se base sur une méthode d'Analyse d'Image multivariée et d'Analyse en Composantes Principales (Wold et al., 1987). Bien que ces méthodes d'analyse de la texture d'images 2D n'aient pas un lien direct avec le traitement de nuages de points, une étape de prétraitement peut permettre le passage d'un objet décrit par un nuage de points en une image 2D, suivie des méthodes décrites plus haut.

## 1.4 Défauts externes du bois à la surface d'un tronc

Les défauts externes du bois jouent un rôle important sur la qualité du bois. Ils ont comme origine des blessures d'exploitation, des fentes causées par le séchage, des fentes dues au gel, des maladies causées par les champignons, des blessures causées par les animaux, et enfin de la ramification due au développement de l'arbre. Parmi les défauts externes, les défauts dus à la ramification de l'arbre sont les plus importants du point de vue de la qualité du bois, car la nodosité diminue les propriétés mécaniques et esthétiques des produits. Dans cette section, nous allons discuter la définition, l'origine et l'impact de ces défauts sur la qualité du bois.

### 1.4.1 Origine de défauts liés à la ramification

La ramification est un phénomène essentiel chez les végétaux qui permet leur développement. Bugnon (1980) a défini la ramification comme "le fait qu'une unité morphologique du corps donne naissance à une ou plusieurs unités nouvelles de même nature fondamentale qu'elle-même. En principe, l'unité génératrice et les unités apparues restent associées en un ensemble dit système ramifié." La ramification a comme origine le bourgeon terminal qui va produire latéralement des bourgeons axillaires que l'on trouve à l'aisselle d'une pièce foliaire, écaïlle ou feuille, lors de l'allongement de chaque pousse annuelle. Selon le moment de sa mise en place, un bourgeon a trois devenir possibles (Colin et al., 2010a) :

- se développer après le repos hivernal pour donner naissance à des rameaux feuillés appelés branches séquentielles,
- mourir par abscission totale ou partielle,
- se maintenir en place sous la forme d'un bourgeon latent.

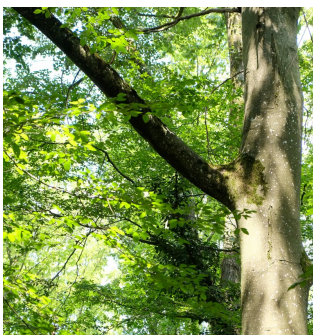
Un bourgeon maintenu en place est également dénommé bourgeon épïcormique. Il peut se développer en un long axe pour former un poil ou une branche épïcormique (gourmand). La présence de traces de branche dans le bois est un défaut appelé noeud qui altère l'orientation du fil du bois, la densité du bois et donc modifie la qualité mécanique du bois. La conséquence du défaut est moins importante s'il reste dans l'aubier, car cette partie est souvent purgée. S'il est présent dans le duramen, l'utilisation pour l'ébénisterie ou pour le tranchage est limitée. Ceci est surtout vrai pour le chêne.



## 1.4.2 Branche séquentielle

**Définition.** Une branche séquentielle est une branche qui se développe à partir d'un bourgeon axillaire, déposé l'année de végétation précédente.

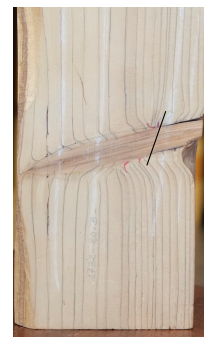
**Impact sur la qualité.** Les branches séquentielles sont mises en place et se développent au fur et à mesure de l'allongement du tronc. La trace d'une branche vivante dans le bois est un nœud adhérent (Figure 1.3 (b)). Pour les espèces qui s'élaguent facilement (chêne), les branches basses les plus âgées meurent, forment des nœuds non-adhérents de longueur réduite puis s'élaguent et disparaissent du tronc en laissant une cicatrice de branche. Pour les espèces qui s'élaguent difficilement (épicéa, sapin), les branches âgées meurent mais restent en place et forment des nœuds non-adhérents de longueur importante (Figure 1.3 (c)). L'impact des branches est donc grand ; leur présence sur le billon n'est pas admise pour un bois de la classe A.



(a) Branche séquentielle



(b) Nœud adhérent

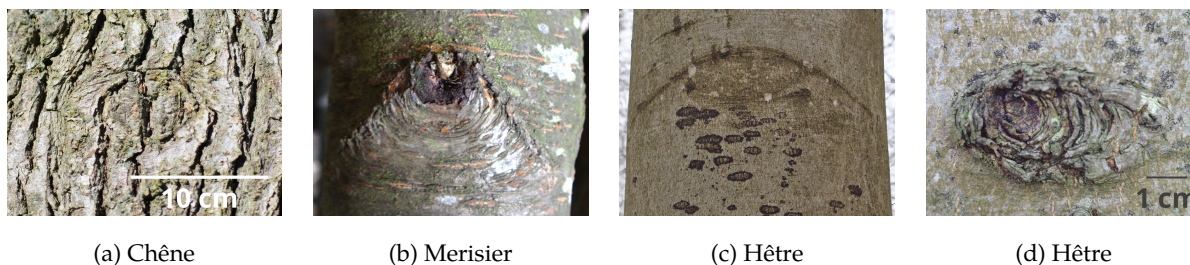


(c) Nœud non-adhérent (à droite de la ligne noire)

**Figure 1.3** Illustration d'une branche séquentielle et des nœuds dans le bois dus aux branches séquentielles.

## 1.4.3 Cicatrice de branche

**Définition.** Pendant le développement de l'arbre, des branches peuvent mourir par un phénomène naturel ou par une opération humaine. Une branche morte laisse sa trace dans le bois (nœud). À la surface, après que la branche soit tombée, la croissance de l'arbre couvre le reste de la branche dans le tronc. Ce recouvrement donne une cicatrice de branche. La Figure 1.4 montre quelques exemples de cicatrices de branches sur les troncs d'espèces différentes. Les formes des cicatrices et les reliefs sont différents selon l'espèce, la condition dans laquelle la cicatrice se forme et l'âge de la cicatrice. Il y a souvent sur les hêtres de grosses cicatrices avec un faible relief (de quelques mm par rapport à la surface de l'écorce) telles que celle représentée dans la Figure 1.4 (c). Détecter ce genre de défauts sur les données dérivées du TLS est un défi important.



**Figure 1.4** Illustrations de cicatrices de branche sur le tronc d'espèces différentes.

**Impact sur la qualité.** L'impact de cicatrice de branche sur la qualité du bois dépend de plusieurs facteurs. Un des facteurs les plus importants est la dimension de la cicatrice : plus la cicatrice est grande, notamment en hauteur, plus la dégradation de la qualité est importante. Sa forme est révélatrice des caractéristiques du nœud à l'intérieur. De plus, l'impact dépend aussi du type de cicatrice. S'il s'agit d'un nœud sain l'impact est moins fort que dans le cas d'un nœud pourri. Un bois de la classe A ne contient pas de cicatrices ou à la rigueur une petite cicatrice.

#### 1.4.4 Amas de bourgeons

**Définition.** C'est un regroupement de moins de 6 bourgeons éplicormiques juxtaposés (le regroupement de 6 ou plus de bourgeons est définie comme un broussin). Les amas proviennent de la multiplication de bourgeons sans que ceux-ci ne se développent en pousse (Colin et al., 2010a).



(a) Un amas de 5 bourgeons (b) Un amas de 2 bourgeons

**Figure 1.5** Illustrations d'amas sur les troncs de chêne.

**Impact sur la qualité.** Conformément à leur définition, ce sont des défauts de petite taille. Leur impact sur la qualité du bois est donc faible et restent plutôt dans l'aubier.

#### 1.4.5 Picot

**Définition.** Ce sont de petits axes courts et épais (jusqu'à 1 cm de diamètre). Ils sont le plus souvent formés suite à l'élagage naturel à répétition de petites pousses éplicormiques feuillées. Le

picot porte des bourgeons, voire des poils, superposés (Colin et al., 2010a). Le picot indique la présence dans la grume de pousses éplicormiques feuillées mortes et donc de petits nœuds noirs ou de traces de bois mort.

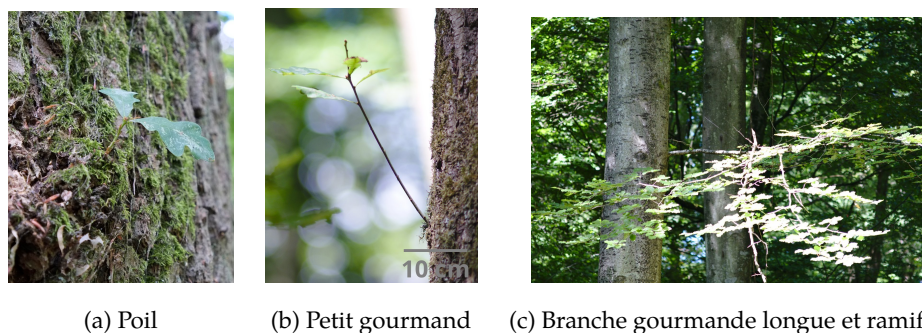


**Figure 1.6** Illustration d'un picot du chêne.

**Impact sur la qualité.** Bien que la taille soit petite avec un diamètre d'environ 1 cm, le picot peut avoir un impact non négligeable sur la qualité du bois en laissant une trace noire dans le bois qui est rédhitoire pour des usages à forte valeur ajoutée tels que le tranchage et la fabrication de merrains. Dans les normes (AFNOR, 2012a; Jourez et al., 2010), la présence de picots est considérée comme un défaut de diamètre de 5 mm. Pour le bois feuillus de la classe A, un seul picot est admis sur 2.5 m de tronc (AFNOR, 2012a).

### 1.4.6 Pousse éplicormique

**Définition.** Les pousses éplicormiques qui incluent les poils, les gourmands et les branches gourmandes sont les pousses issues d'un bourgeon éplicormique. Le poil est âgé d'au maximum 2 ans, n'est pas ramifié, et mesure environ 5 cm de long. Le gourmand est plus long, ramifié et âgé de plusieurs années. La branche gourmande est un gourmand de grande dimension comparable à celle d'une branche séquentielle bien développée.



(a) Poil

(b) Petit gourmand

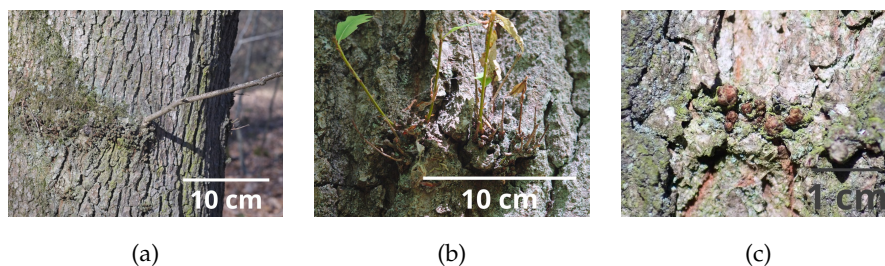
(c) Branche gourmande longue et ramifiée

**Figure 1.7** Illustrations de pousses éplicormiques.

**Impact sur la qualité.** Les gourmands présents sur le tronc constituent l'un des défauts les plus fréquents. Leur impact sur la qualité du bois dépend de la taille de la trace laissée dans le bois. Si la trace est petite et faible en couleur, la dégradation est négligeable. C'est le cas des poils. Par contre, si la trace est assez grande pour altérer le fil du bois et/ou présente une couleur foncée, l'impact sur la qualité est considérable (Fontaine et al., 2002). C'est le cas des branches épicorniques bien développées.

### 1.4.7 Broussin

**Définition.** C'est une singularité formée d'un regroupement de bourgeons, d'amas, de poils, voire de picots et des restes de rameaux, tous juxtaposés, pouvant constituer une véritable protubérance sur l'écorce. La présence de cette singularité aurait un déterminisme génétique, mais rien ne le prouve de manière irréfutable (Colin et al., 2010a).



**Figure 1.8** Illustrations de broussins sur les troncs de chênes : (a) un broussin composé d'un gourmand et des bourgeons, (b) un broussin composé de poils et de bourgeons et (c) un broussin composé uniquement de bourgeons.

**Impact sur la qualité.** Conformément à la définition des broussins, ils peuvent avoir une taille et une forme très variées (Figure 1.8). L'impact sur la qualité du bois dépend donc de la dimension et de la composition du broussin. La trace laissée dans le bois du broussin est souvent appelée "patte de chat" (Figure 1.9). C'est un défaut très pénalisant.



**Figure 1.9** Illustration des traces d'un broussin : "patte de chat".

### 1.4.8 Sphéroblaste

**Définition.** Un sphéroblaste est un bourgeon dont la base produit du xylème qui recouvre progressivement le méristème apical du bourgeon.



**Figure 1.10** Illustration d'un sphéroblaste.

**Impact sur la qualité.** Les sphéroblastes se présentent principalement sur le hêtre et n'ont aucun impact sur la qualité du bois (Colin et al., 2007).

## 1.5 Objectifs de la thèse

L'objectif principal de la thèse est de développer des algorithmes afin de valoriser les données LiDAR terrestre pour quantifier les défauts externes de grumes sur pied ou de bois ronds. Plus précisément, il s'agit d'analyser la rugosité de la surface de l'écorce, pour détecter les défauts externes. A ce stade on se place dans les conditions de très haute résolution spatiale des données 3D afin que la résolution spatiale permette de décrire les plus petits défauts. Ensuite, il s'agit d'identifier le type de défaut et mesurer ses dimensions et position. La méthode recherchée se veut générique, fiable et robuste sur le plan des espèces et des types de défauts à traiter. De plus, le traitement de données 3D à haute densité exige que les algorithmes soient efficaces en terme de temps de calcul et de gestion de la mémoire utilisée.

Nous proposons donc une approche décomposée en 3 étapes. Première étape : détection de zones susceptibles de correspondre aux défauts. Deuxième étape : identification du type de défaut pour chaque zone détectée . Troisième étape : caractérisation des défauts (Figure 1.11). A chaque étape, des expérimentations ont été menées pour valider l'efficacité de notre approche. Le matériel végétal utilisé dans ces expérimentations est résumé dans la Figure 1.12.

L'objectif de la première étape est de repérer rapidement les zones sur la surface où une analyse du relief révèle une singularité susceptible de correspondre à un défaut. La méthode proposée et le résultat des expérimentations seront présentés dans le chapitre 2.

La deuxième étape consiste à identifier le type de défaut de chaque zone détectée lors de la première étape. Cette étape a deux objectifs principaux. Le premier objectif est de confirmer la pertinence de la détection et d'enlever les faux positifs (les zones ne correspondant à aucun défaut) et/ou de détecter si une zone correspond à plusieurs défauts. Le deuxième objectif est de classer les défauts selon leur origine biologique définissant leur type. Une approche basée sur l'apprentissage automatique est proposée et sera présentée dans le chapitre 3.

La troisième étape a pour but de déduire des caractéristiques des défauts telles que la position et les dimensions. Ensuite, une utilisation de ces informations en vue de classer les bois est aussi présentée dans le chapitre 4.

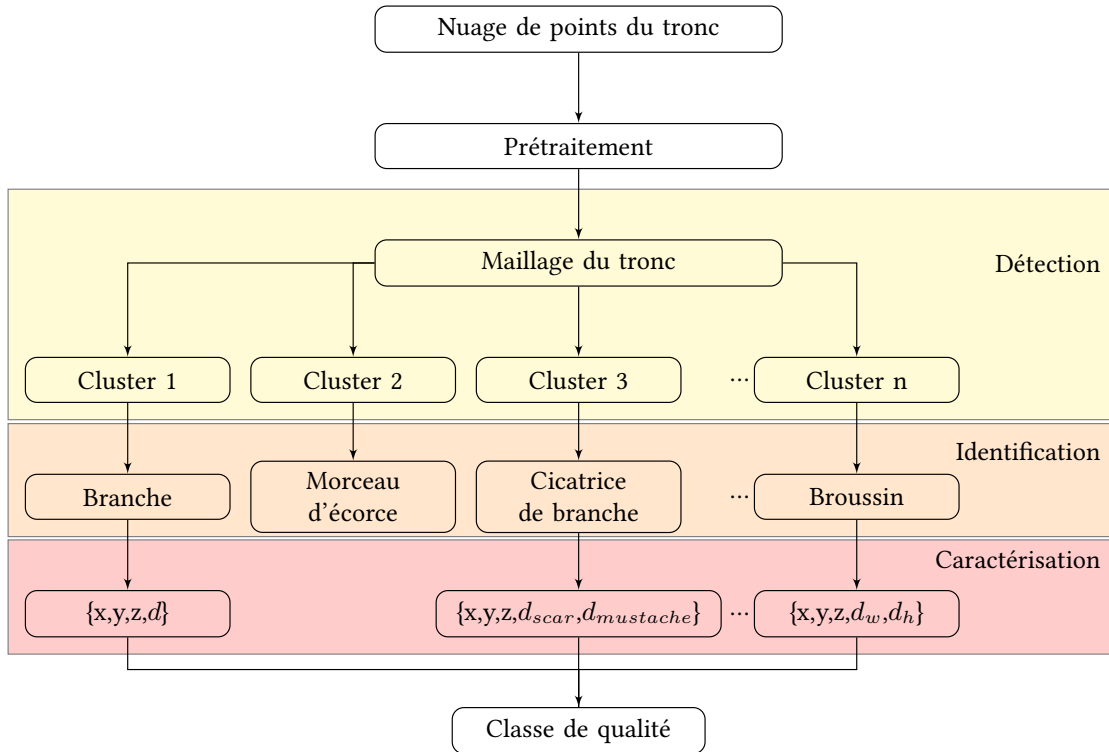


Figure 1.11 Schéma de l'approche en 3 étapes.

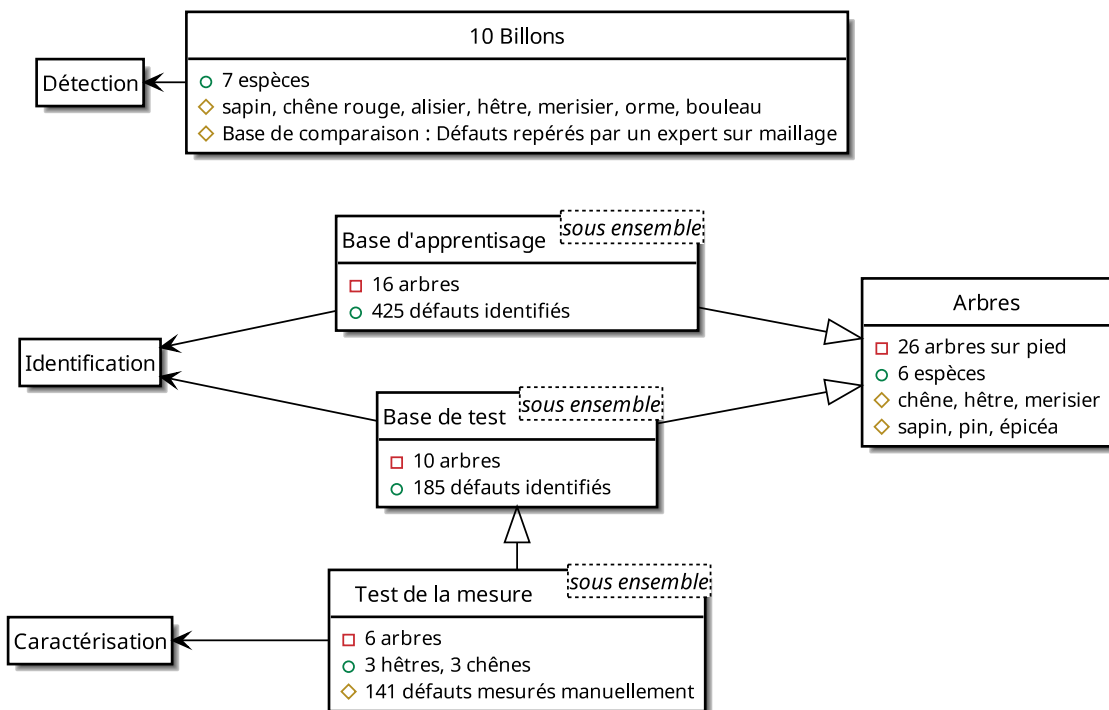


Figure 1.12 Description du matériel végétal utilisé dans les expérimentations à chaque étape.





## Chapitre 2

# Détection des défauts à la surface du tronc

### 2.1 Contexte et objectifs

Les données 3D à forte densité issues d'un Lidar terrestre ou d'un cubeur contiennent une information qui peut potentiellement servir au classement des bois ronds et à localiser les défauts. Ainsi, cela apparaît comme une alternative à la méthode de référence dans ce domaine qui est la tomographie à rayons X notamment pour le classement de la qualité du bois rond, voire l'optimisation de sa transformation avec moins de contraintes réglementaires et un coût d'investissement moins élevé. Dans le domaine forestier, le LiDAR terrestre offre la possibilité de scanner des arbres sur pied, et d'évaluer leur qualité en détectant les défauts locaux sur les surfaces des troncs.

Au cours de la dernière décennie, plusieurs travaux de recherche ont montré que les corrélations existant entre les défauts internes et externes atteignent des coefficients de détermination ( $R^2$ ) supérieurs à 0,6 (Thomas, 2009; Stängle et al., 2013; Racko, 2013; Pyörälä et al., 2017). Ces corrélations élevées, implicites dans les normes de classement des bois ronds, confirment que les défauts externes peuvent être utilisés pour déduire l'information sur les défauts internes. Cependant, les recherches sur la segmentation des défauts sur les surfaces du tronc sont encore peu nombreuses (Schütt et al., 2004; Thomas et al., 2007; Kretschmer et al., 2013; Stängle et al., 2013). De plus, ces recherches se sont concentrées sur de larges défauts (Thomas et al., 2007) ou sont semi-automatiques (Kretschmer et al., 2013; Stängle et al., 2013; Schütt et al., 2004). Pour répondre à l'ambition énoncée précédemment, il est nécessaire de détecter automatiquement tous les types de défauts externes liés à la ramification comme cible principale.

L'objectif de cette première partie est de repérer rapidement les zones sur la surface où une analyse du relief révèle une singularité susceptible de correspondre à un défaut. À partir des

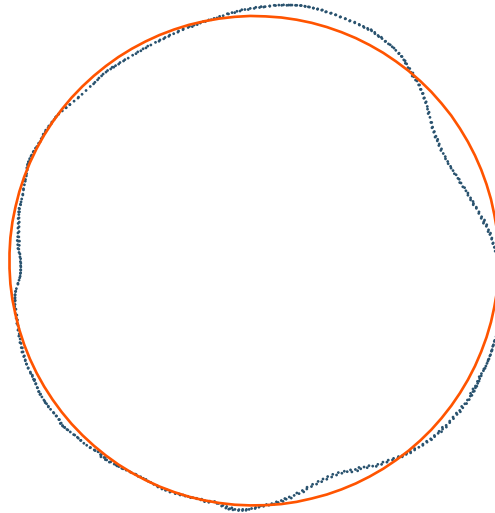
données 3D représentant la surface du tronc, l'idée principale est d'analyser la variation du relief de la surface qui peut être évaluée à partir de la distance de chaque point du nuage de points à une référence. Morphologiquement, les défauts sur le tronc peuvent être classés en deux catégories selon leur relief par rapport la surface du tronc. La première catégorie comprend les défauts avec un très faible relief tels que la cicatrice de branche, le picot, le petit sphéroblaste et l'amas de bourgeons. La seconde comprend les défauts qui sont nettement en relief par rapport au tronc tels que la branche, le broussin et le gourmand. La différence morphologique entre les deux catégories de défaut exige des approches différentes pour les segmenter. La segmentation de défauts de la première catégorie est présentée dans le premier article (section 2.5). Les détails des algorithmes et des discussions sur l'influence des paramètres principaux sont présentés dans le deuxième article (section 2.6). La segmentation de défauts de la deuxième catégorie et des améliorations pour adapter le traitement au volume des données 3D représentant des grumes réelles et pour détecter des défauts de première catégorie sont présentées dans le troisième article (section 2.7).

## 2.2 Méthodologie

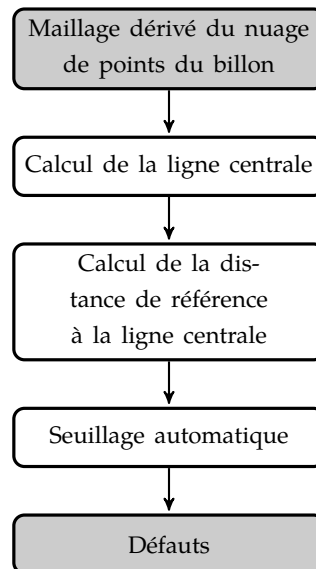
Nous pouvons détecter les défauts par les changements de relief sur l'écorce du tronc. En effet, les défauts correspondent souvent aux régions avec un relief plus fort (Carpenter et Jones, 1989). Des travaux existants ont détecté ces régions en comparant la surface du tronc avec une surface de référence. La surface de référence a souvent été définie par des cylindres (Schütt et al., 2004; Stängle et al., 2013; Kretschmer et al., 2013) ou par des cercles (Thomas et al., 2007; Thomas et Mili, 2007). A cause des irrégularités de la forme du tronc, la surface de référence créée par les formes primitives ne se rapproche pas toujours bien de la surface réelle du tronc surtout lorsque celui-ci présente une forme irrégulière (voir la Figure 2.1 pour exemple) et par conséquent, on ne peut pas détecter de petits défauts. En visant une méthode générique qui fonctionne avec des espèces différentes et tous les types de défauts liés à la ramification, nous proposons une nouvelle approche se basant sur l'estimation de la distance de référence à la ligne centrale pour chaque point du nuage de points à partir d'un patch contenant ses points voisins. La Figure 2.2 résume notre approche pour détecter les défauts avec un faible relief. La segmentation des branches sera présentée dans la section 2.2.4.

### 2.2.1 Calcul de la ligne centrale

Le principe du calcul de la ligne centrale du tronc est basé sur la concordance des zones de l'espace 3D visées par les vecteurs normaux en chaque point du nuage de points. Le vecteur normal d'un point est calculé à partir d'un ajustement de plan à ses points voisins. Ainsi, nous balayons l'espace dans la direction du vecteur normal du point et sur une longueur supérieure au rayon du tronc (orienté vers l'intérieur de l'objet). Pendant le scan, nous augmentons d'une unité le compteur d'accumulation (score d'accumulation) de chaque voxel de l'espace digitalisé (image



**Figure 2.1** Exemple de la difficulté d’ajustement d’un cercle à la section transversale d’un tronc avec une forme irrégulière.



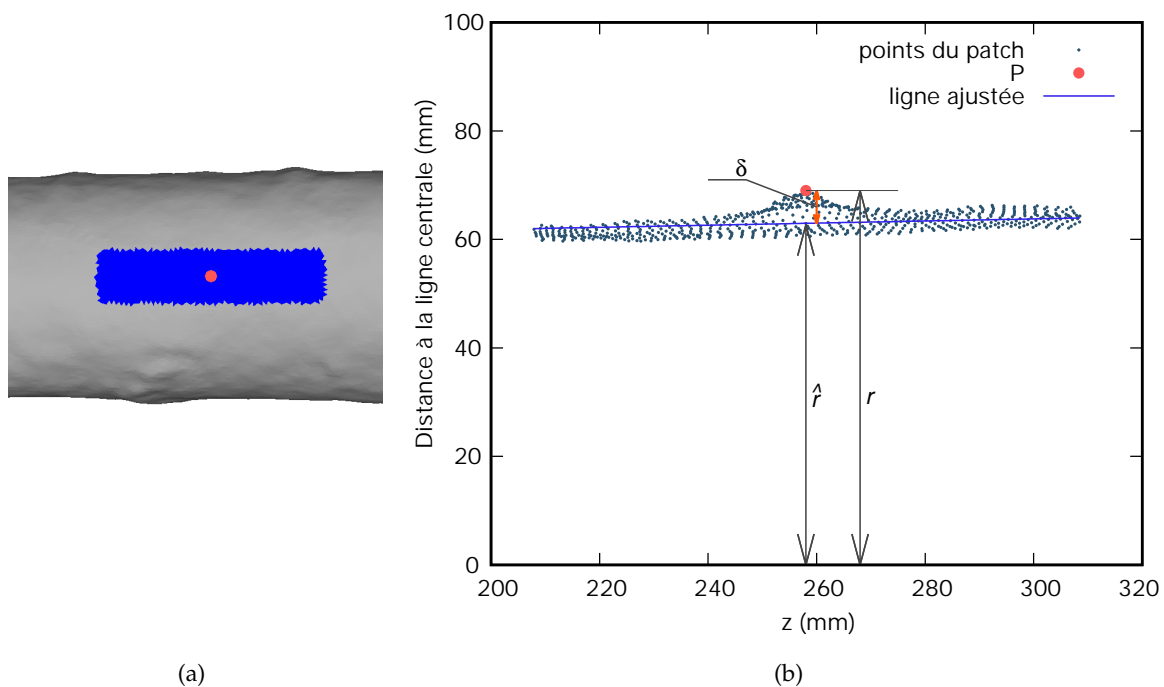
**Figure 2.2** Schéma pour la segmentation des défauts

d’accumulation) traversé par le scan. Le principe du calcul de la ligne centrale est que dans un objet tubulaire tel qu’un tronc, les vecteurs normaux convergent vers son axe médian. La première version du calcul de la ligne centrale a été basée sur la recherche des maxima locaux dans l’image d’accumulation (Kerautret et al., 2015), suivie par une étape d’optimisation. Les principes de cette méthode seront détaillés dans le premier article (section 2.5) et les descriptions d’implémentation seront décrites dans le deuxième article (section 2.6). Une amélioration de cette méthode sera présentée dans le troisième article (section 2.7) en vue de prendre en compte des formes de troncs plus complexes que la première version ne supporte pas. Les améliorations portent aussi sur la capacité à traiter des données volumineuses associées à un tronc complet scanné à haute résolution et à la présence des branches. Dans cette article nous présenterons aussi une amélioration du calcul

de la ligne centrale en se basant sur la nouvelle méthode de l'accumulation des vecteurs normaux (Kerautret et al., 2016).

## 2.2.2 Calcul de la distance de référence à la ligne centrale

Nous définissons la distance de référence à la ligne centrale d'un point comme la distance entre le point et la ligne centrale d'un billon idéalisé qui n'a pas de défaut. Selon notre expertise de l'origine biologique des défauts, la plupart des points, dans un patch de forme allongée dans le sens vertical de l'arbre ne font pas partie d'un défaut. À partir de cette observation, la distance de référence à la ligne centrale d'un point  $P$  est calculée par une analyse de la distribution de la distance à la ligne centrale des points appartenant au patch centré sur  $P$  (voir la Figure 2.3). La structure de données "K-d tree" (Bentley, 1975) est utilisée pour accélérer la recherche des points voisins. Dans le troisième article (section 2.7) nous proposons des adaptations aux données volumineuses en utilisant une technique de sous-échantillonnage du nuage de points.



**Figure 2.3** Un patch longitudinal du point  $P$  (a) et l'ajustement d'une ligne au patch pour calculer la distance de référence à la ligne centrale ( $\hat{r}$ ) (b). L'axe  $z$  représente la position longitudinale dans une orientation globale du billon selon sa longueur.

## 2.2.3 Seuillage automatique

La méthode de seuillage automatique de Rosin (Rosin, 2001) permet de classer le nuage de points à partir de l'histogramme de la différence entre la distance à la ligne centrale et la distance

de référence à la ligne centrale (la valeur du  $\delta$  dans la Figure 2.3 (b)). Le détail de l’algorithme sera présenté dans le premier article 2.5. Le deuxième article 2.6 présentera les détails techniques et une implémentation de l’algorithme.

## 2.2.4 Segmentation des branches et des défauts présentant un axe long

La méthode présentée précédemment est bien adaptée aux défauts de faible relief. Elle nécessite de travailler sur un nuage de points où les branches sont exclues. Cette section présente une proposition pour segmenter les branches sachant que des méthodes de segmentation de branches existent dans les travaux dédiés à la reconstruction géométrique d’arbres (Côté et al., 2011; Hakenberg et al., 2015; Raumonon et al., 2013). Dans la version finale de notre outil complet de caractérisation des défauts externes, cette segmentation est réalisée avant la détection de petits défauts. Les branches sont segmentées par la méthode suivante (section 2.7). L’idée principale est de séparer le nuage de points en deux ensembles disjoints dont les points appartiennent au tronc ( $\mathcal{T}$ ) et aux branches ( $\mathcal{B}$ ) respectivement. Les étapes sont les suivantes :

- Estimation du rayon moyen  $r_m$  du tronc en utilisant le mode (la valeur la plus fréquente) de la distance à la ligne centrale de tous les points du nuage de points.
- Division du volume du nuage de points en des secteurs angulaires cylindriques dont la ligne centrale est leur centre. La hauteur et l’angle de secteur sont respectivement de  $l$  mm et de  $\frac{l}{r_m}$  radian respectivement. Dans notre expérimentation, la bonne valeur pour le paramètre  $l$  était entre 50 mm et 100 mm ce qui permet à la fois de bien segmenter les branches et de garantir la performance de l’algorithme de segmentation de petits défauts.
- Pour chaque secteur angulaire cylindrique, le point le plus proche au centre du tronc est rajouté dans l’ensemble  $\mathcal{T}$ , les autres points de la portion sont ajoutés à l’ensemble  $\mathcal{B}$ .
- Pour chaque point  $P$  de l’ensemble  $\mathcal{T}$ , nous cherchons le sous-ensemble  $\mathcal{S}$  de l’ensemble  $\mathcal{B}$ , tel que la distance entre un point  $S_i \in \mathcal{S}$  et  $P$  soit inférieure ou égale à  $\sqrt{2}l$ . Enfin, nous déplaçons le sous-ensemble  $\mathcal{S}$  de  $\mathcal{B}$  dans l’ensemble  $\mathcal{T}$ .
- Le résultat de la procédure est deux ensembles disjoints : l’ensemble  $\mathcal{B}$  contenant les points qui appartiennent aux branches et l’ensemble  $\mathcal{T}$  contenant les points qui appartiennent au tronc.

## 2.3 Résultats principaux

Les résultats du premier test (section 2.5) sur 10 billons de 7 espèces différentes telles que orme, sapin, merisier, hêtre et chêne rouge ont montré que notre méthode peut segmenter les défauts avec précision et semble être robuste pour les différentes espèces d’arbres et caractéristiques

géométriques. Nous avons comparé ces résultats avec des données de référence (vérité terrain) au niveau des points. Le critère choisi pour évaluer la performance de la segmentation est la mesure  $F_1$  qui est la moyenne harmonique de la précision et du rappel, deux variables calculées à partir des nombres de faux et vrais positifs. La valeur de la mesure  $F_1$  moyenne était de 0,71. Ceci indique qu'il y a une forte correspondance entre les points détectés et les points de référence. Nous avons refait cette expérimentation dans le troisième article pour vérifier nos améliorations, la nouvelle mesure  $F_1$  était 0,76. Cette valeur de  $F_1$  était supérieure à celle d'une approche d'ajustement des cylindres qui était de 0,38. Cependant, les résultats contiennent un certain nombre de faux positifs (fausses détections) qui seront traités par une analyse plus fine dans le chapitre suivant.

La segmentation des branches a été testée dans le troisième article 2.7. Nous avons testé notre segmentation dans des cas difficiles tels qu'un tronc flexueux avec un nombre important de branches et une seule face du tronc. Dans tous les cas notre algorithme a détecté 33 branches sur 2 troncs avec une exactitude de 100%.

## 2.4 Conclusion

Ce chapitre a proposé une nouvelle approche pour détecter les défauts sur la surface des troncs d'arbre à partir de données TLS. La méthode proposée calcule pour chaque point du nuage de points un patch des points voisins pour estimer une distance de référence à la ligne centrale du tronc. Les résultats des expérimentations ont montré que notre méthode peut segmenter avec précision les défauts ayant une forme complexe comme les "moustaches de chinois" et les petits défauts. Dans un premier temps, notre méthode a été conçue pour détecter des défauts à faible relief tels que les cicatrices de branches, les amas et les broussins (qui ne se composent que de bourgeons et de poils). Des améliorations ont été apportées pour que la méthode puisse détecter tous les défauts liés à la ramification y compris les branches. Des tests réalisés avec des troncs de différentes espèces telles que chêne, chêne rouge, sapin, hêtre et merisier ont montré que la méthode finale répondait bien à la présence de branches, la courbure du tronc, la rugosité de l'écorce et l'irrégularité de la forme des sections transversales et le manque de données sur toute la périphérie du tronc. De plus, en comparant avec une méthode basée sur un ajustement cylindrique pour la surface de référence, notre méthode s'est avérée plus précise et robuste. Le résultat de la méthode présentée peut être utilisé pour caractériser les défauts externes d'une grume ou d'un tronc d'arbre sur pied afin d'évaluer la qualité de son bois.

Le code concernant le premier article (section 2.5) est disponible sur l'entrepôt du github<sup>1</sup>. Une démonstration en ligne de la segmentation de défauts est aussi disponible<sup>2</sup> qui permet de reproduire le résultat de la segmentation des 10 billons. Le troisième article (section 2.7) sera potentiellement publié dans le journal IPOL (Image Processing On Line).

1. <https://github.com/vanthongyen/treelogdefectsegmentation>

2. [http://ipol-geometry.loria.fr/~kerautre/ipol\\_demo/TrunkDefaultMeasure](http://ipol-geometry.loria.fr/~kerautre/ipol_demo/TrunkDefaultMeasure)

## **2.5 Article 1 : Segmentation of defects on log surface from terrestrial Lidar data**

Publié dans les actes de la conférence Pattern Recognition (ICPR), 2016 23rd International Conference on, IEEE, pp 3168–3173

# Segmentation of defects on log surface from terrestrial Lidar data

Van-Tho Nguyen <sup>\*</sup>, Bertrand Kerautret <sup>†</sup>, Isabelle Debled-Rennesson <sup>†</sup>, Francis Colin <sup>\*</sup>,  
Alexandre Piboule<sup>‡</sup>, Thiéry Constant <sup>\*</sup>

<sup>\*</sup> LERFOB, AgroParisTech, INRA, 54000 NANCY France.

<sup>†</sup> LORIA, UMR CNRS 7503, Université de Lorraine, 54506 Vandœuvre-lès-Nancy, France.

<sup>‡</sup> ONF, RDI, 5 Rue Girardet, 54000 Nancy, France.

**Abstract**—Segmentation of defects on the tree log surface remains a challenge due to the unclear separation between the foreground and the background and the high variability of the tree surface. Even if some first works exist to process specific tree species, a generic method robust to various species is missing. We propose a new approach for segmenting defects on log surface based on the tabular object analysis. We firstly compute the log centerline by surface normal accumulation and then threshold the point cloud by the difference between the distance to the centerline and the reference distance estimated from a patch of neighbors. The performance of the proposed approach was experimented and compared on ten logs recovered from different species. The results showed that our approach outperformed other method based on cylinder detection and was robust to several tree species. The results can be reproduced and compared on an online demonstration.

## I. INTRODUCTION

The segmentation and the analysis of tubular objects are recurrent problems in the domain of computer imagery. They are often motivated by applications coming from various fields like medicine, biology or industry [1], [2]. By nature, the tubular shapes are naturally present in medical imagery with for instance the blood vessel segmentation [3], [4] and analysis [5]. Always in the biology field, they are also present within 3D volumic imagery of tree log with the knot segmentation and analysis [6]. Such analysis allows the users to recover and measure the knot orientation and volume, useful to industrial process with the sawing pattern optimization or to agronomic researchers [7]. Since the 3D volumic data are acquired from medical CT scanner, these analyses are not applicable to measure wood quality on standing tree. To overcome this limitation, we focus on data coming from Terrestrial Laser Scanning (TLS) which is increasingly used for characterizing standing trees in forestry. This technology provides high resolution 3D point cloud with millimeters accuracy. In particular, this work aims to detect defects on the bark surface. These defects are in general associated with more or less protruding regions resulting from the inclusion of the internal defect by the tree growth; in fact the development of branches and epicormic shoots form knots in wood. On the surface, these protruding defects have various forms like the "Chinese mustache" or an ellipse shape (Figure 1). Identifying these defects is an important task for assessing wood quality in forest inventory, or in sawmill yard, since

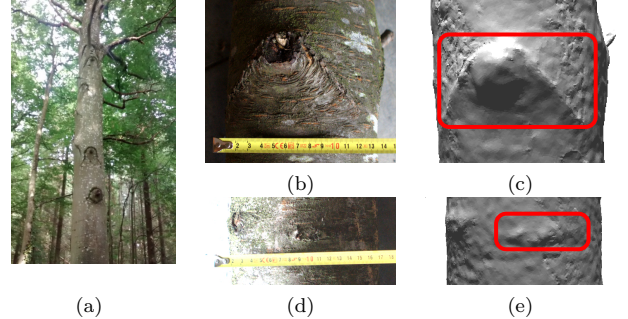


Figure 1. (a) Defects on a standing tree in forest. (b,c) A Chinese mustache and its corresponding mesh. (d,e) A small defect in ellipse shape and its corresponding mesh.

internal wood knots have a structural link with defects on the surface[8], [9], [10]. However, due to the complex shape of log like curvature, taper, ovality or even more complex (see Figure 2(a)) and the roughness of the bark, the automatic segmentation of defects on the log surface is not trivial, especially when it is intended to isolate the small or tiny defects (defects which rise a little from normal surface). The existing works on identification of external defects from TLS data are mostly semi-automatic. Schütt et al [11] define a region based segmentation of gray scale image derived from the distance between 3D point cloud of the log and the fitted cylinders. Then, the user can interactively decide whether a region is a defect or not. Using a similar cylinder fitting method, [9], [12] used a 3D visualization of this distance to locate and measure defect sizes on log surface. The severe defects with height more than 2.2 cm and diameter more than 7.5 cm can be detected by a contour analysis of 2D image [13], [14]. The 2D image is constructed by the distance between point cloud and reference circles.

In this context, the main purpose of this work was to segment wood defects detectable on log surface from TLS data. The proposed method is based on a method assessing the reference distance to the centerline for each 3D point taking the neighboring points into account. This method was expected to work with different logs without concern of the species, bark roughness, longitudinal flexuosity or cross-sectional ovality. The following section describes the new method to segment defects on log surface. The last section shows the agreement with delineation of surface



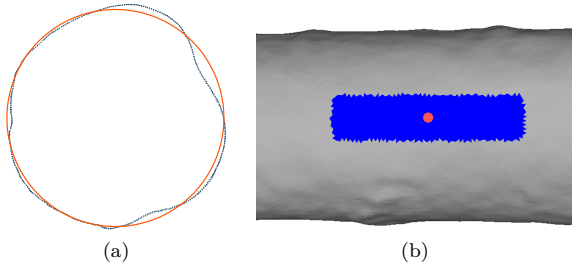


Figure 2. (a) Example of a fitted circle to a log section without defects. (b) A patch in blue is used to estimate reference distance to the centerline for orange point.

defects made by an expert.

## II. METHODOLOGY

The main idea to segment the defects on log surface was to define a reference surface that approximates the defect-free part of the log surface. The points on the log surface can be then classified by their distance to the reference surface. The reference surface was often defined by cylinders [11], [9], [12] or by circles [13], [15]. However, due to the complex shape of the log, these primitive shapes did not always well fit. Figure 2(a) shows the fitted circle on a log section without defects. We can see that the circle does not fit well the data points, with some parts of the data may be recognized as defects. To deal with such frequent cases, we proposed a novel method to estimate the reference distance to the centerline. The reference distance to the centerline here represents the distance from the point to the centerline of an idealized log which would have no defects. Based on our expertise in the biological origin of these defects, most points in the longitudinal patch (see Figure 2 (b)) were considered not belonging to a defect. Hence, the reference distance to the centerline of a point can be deduced from the analysis of the distance to the centerline of the points belonging to the patch (see in the section II-C).

In this section, we propose a method that firstly computes the centerline of the log and then, we transform the point cloud to cylindrical coordinate system to simplify the neighbors search. We then compute a reference distance to the centerline for each point from its surrounding patch. Lastly, an automatic thresholding method is used to classify the point cloud base on the difference between the distance to the centerline and the reference distance to the centerline (see Figure 3).

### A. Compute centerline

In the proposed method, the estimation of the log centerline is an important step since it can differ from a simple straight line (see for instance Figure 5). To apply a precise and simple centerline extraction, we exploit the algorithm based on surface normal accumulation recently proposed [1].

The main idea of this simple method to implement, is illustrated on Figure 4. The first step consists in starting from a mesh face  $f_i$  with its normal vector  $\vec{n}_i$  and applying

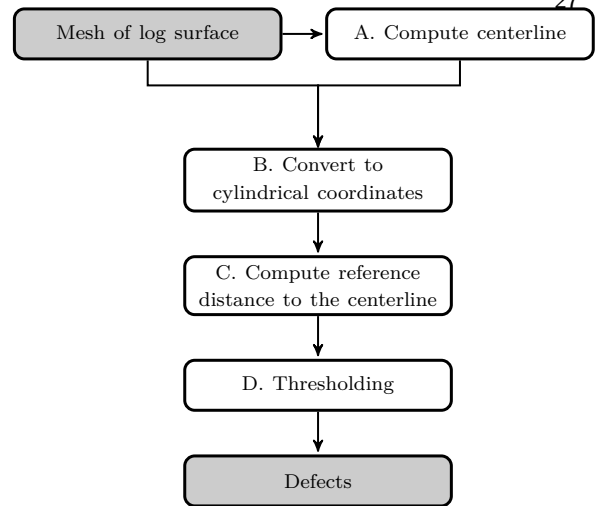


Figure 3. Overview of the algorithm

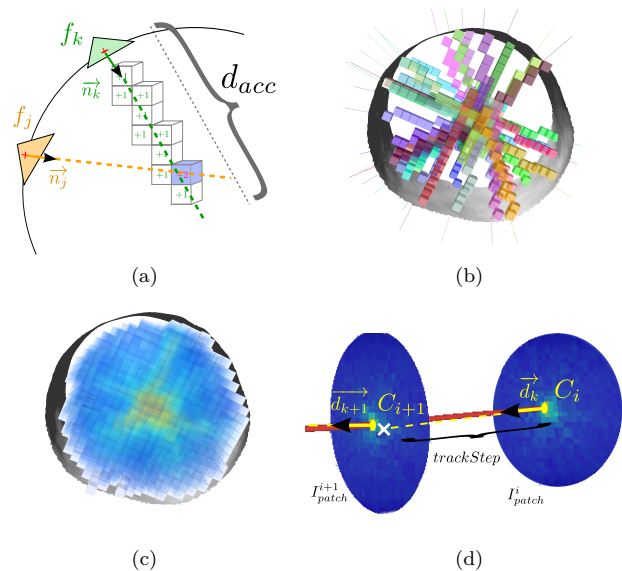


Figure 4. Illustrations of the main ideas of the centerline method based on surface normal accumulation: (a,b) scan from mesh faces along a distance  $d_{acc}$ , (c) accumulation score, (d) tracking centerline points  $C_i$  in the direction  $\vec{d}_k$ .

a normal oriented directional scan along a distance  $d_{acc}$  (oriented toward the object interior). During this scan, we increase by one all voxels of the digitized space which are crossed by the 3D line defined from the face center and the direction  $\vec{n}_i$  (see Figure 4 (a,b)). From this process, we obtain an accumulation map indicating for each voxel, what is the probability to be in a centerline region (Figure 4(c)). Based on this map, the centerline curve is obtained by a simple tracking process by looking at the local maxima accumulation points  $C_i$  of 2D image patch resulting from the projection in the  $\vec{d}_k$  direction (image d). This direction  $\vec{d}_k$  is computed from all the points which are at the origin of the accumulation value (see [1] for details). Due to the digital space or noise the resulting centerline can present

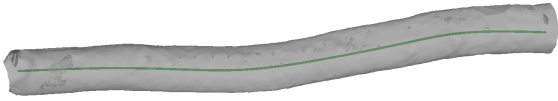


Figure 5. Center line of a Beech log

some irregularities and the method propose an elastic force optimization process to obtain a smooth curve.

Obtaining a smooth centerline curve is an important requirement to our method, however the optimization process is efficient only if the tubular shape has a constant diameter. In our context such condition is not insured and may depend on the tree species. To overcome this limitation, we have used another smoothing process based on cubic spline which is implemented in GNU scientific library [16]. This interpolation produces a smooth polyline with a small segments and Figure 5 shows the final result of a computed centerline.

### B. Convert to cylindrical coordinates

The point cloud is represented in Cartesian coordinate system which is not optimal for the log surface which has a cylindrical-like shape. We thus convert the point cloud to a cylindrical coordinate system. A point  $P(x, y, z)$  in Cartesian coordinate system is converted to a point  $\rho(r, \theta, z)$  in cylindrical coordinate system with:

- $r$  is the distance from  $P$  to the centerline (see Figure 6).
- $z$  is the height of the point along the centerline (see equation 4 and Figure 6).
- $\theta$  is the angle between the radial vector and the  $Oy$  axis of local coordinates of the segment.

The centerline is a polyline represented by  $m + 1$  points  $\mathcal{C} = \{C_0, C_1, \dots, C_m\}$  which form  $m$  segments  $S = \{s_0, s_1, \dots, s_{m-1}\}$  with  $s_i = C_i C_{i+1}$  and  $\vec{s}_i = \overrightarrow{C_i C_{i+1}}$ . To convert the point cloud to cylindrical coordinate system we define the local coordinate system corresponding to each segment  $s_i$ . Each point in the point cloud is then assigned to one of these local coordinate systems. The following is the detail of the algorithm.

- 1) **Define local coordinates.** For each segment  $s_i$ , we define a local coordinate system  $O_i x_i y_i z_i$  where  $\{\vec{u}_i, \vec{v}_i, \vec{w}_i\}$  are basis vectors and the origin  $O_i$  is  $C_i$ .  $\vec{w}_i$  is the normalized vector of  $\vec{s}_i$ .  $\vec{v}_i$  is recursively computed (see Figure 6) as follows. Let  $\vec{n}_i$  be the normal vector of the plane formed by  $\vec{w}_i$  and  $\vec{v}_{i-1}$ :  $\vec{n}_i = \vec{w}_i \times \vec{v}_{i-1}$ ;  $i \in [1, m-1]$ .  $\vec{v}_i$  is the cross product of  $\vec{w}_i$  and  $\vec{n}_i$ . In case  $i = 0$ , we define  $\vec{v}_{i-1} = \vec{v}$ , the unit vector of  $Oy$  in the global coordinates. And finally,  $\vec{u}_i$  is computed by the cross product of  $\vec{v}_i$  and  $\vec{w}_i$ :  $\vec{u}_i = \vec{v}_i \times \vec{w}_i$ .
- 2) **Compute cylindrical coordinates.** Given a point  $P$ , we need to determine which local coordinate system that  $P$  belongs to. The domain of the local coordinate system  $O_i x_i y_i z_i$  is defined by two planes  $m_i$  and  $m_{i+1}$  such that the angle between

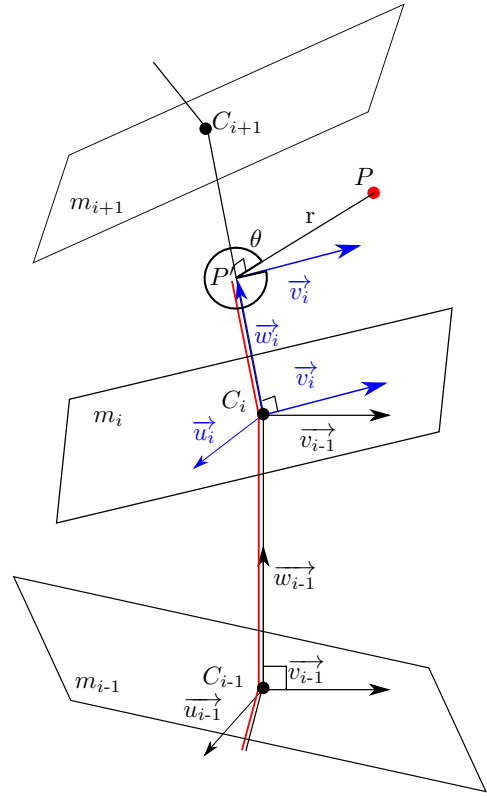


Figure 6. Computation of the coordinate system  $\rho(r, \theta, z)$  from the point  $P$  with  $z$  is the red line.

$m_i$  and  $-\vec{s}_{i-1}$  equals the angle between  $m_i$  and  $\vec{s}_i$ . A point  $P$  belongs to  $O_i x_i y_i z_i$  if and only if  $P$  is located inside the domain of  $O_i x_i y_i z_i$  (see Figure 7). Supposing that  $P$  belongs to  $O_i x_i y_i z_i$ , we now can calculate the radial coordinate  $r_p$ .

$$r_p = \|\overrightarrow{P'P}\| \quad (1)$$

with  $P'$  is the orthogonal projection of  $P$  on  $s_i$ . The azimuthal coordinate  $\theta_p$  equals the angle between  $\overrightarrow{P'P}$  and  $\vec{v}_i$ .

$$\theta_p = a \cos\left(\frac{\overrightarrow{P'P} \cdot \vec{v}_i}{\|\overrightarrow{P'P}\|}\right) \quad (2)$$

Since the equation 2 returns a value in  $[0, \pi]$  even when the angle between  $\overrightarrow{P'P}$  and  $\vec{v}_i$  is greater than  $\pi$ . This is the case that  $P$  is on the left side of the plane  $O_i y_i z_i$ , or in other words, the dot product of  $\overrightarrow{P'P}$  and  $\vec{u}_i$  is negative, an adjustment is required:

$$\theta_p = 2\pi - \theta_p \quad (3)$$

The vertical coordinate is computed as follows:

$$z_p = \sum_{j=0}^{i-1} \|\vec{s}_j\| + \|\overrightarrow{C_i P} \cdot \vec{s}_i\| \quad (4)$$

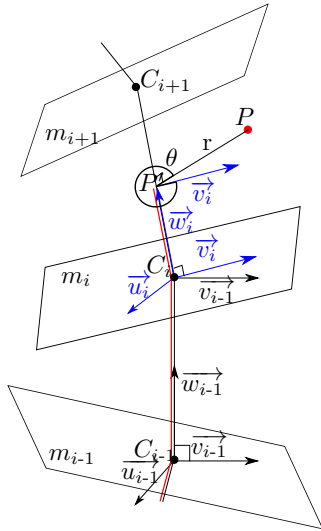


Figure 7. The domain of the local coordinate system  $O_i x_i y_i z_i$  formed by two planes  $m_i$  and  $m_{i+1}$ .

### C. Compute reference distance to the centerline

In the cylindrical coordinate system, at the point  $\rho(r, \theta, z)$  we compute a narrow patch  $\mathcal{P}_\rho$  containing the neighbors of  $\rho$ .

$$\mathcal{P}_\rho = \{\rho_j \mid |\theta_{\rho_j} - \theta_\rho| \leq \frac{\varphi}{2}, |z_{\rho_j} - z_\rho| \leq \frac{\tau}{2}\} \quad (5)$$

with  $\varphi$  and  $\tau$  the parameters determining the patch size. We hypothesized that in a narrow patch, two points with the same vertical coordinate have the same radial coordinate regardless of their azimuthal coordinate. Hence, the reference distance to the centerline  $\hat{r}$  of  $\rho$  only depends on the vertical  $z$ . Moreover, the conversion step from Cartesian coordinates to cylindrical coordinates removed log flexuosity. Thus, the relation between  $\hat{r}$  and  $z$  is linear and can be represented by a straight line (see Figure 8). Since some big defect can perturb the fitting process, the points with radial coordinate greater than the mean plus 2 standard deviations are eliminated before carrying out the least square fitting.

$$\hat{r} = az + b \quad (6)$$

The reference distance to the centerline of the point  $\rho$  is obtained by the equation 6.

$$\hat{r}_\rho = az_\rho + b \quad (7)$$

### D. Thresholding

Firstly, we compute the difference between the distance to the centerline  $r$  and the reference distance to the centerline  $\hat{r}$  for all points:  $\delta_\rho = r_\rho - \hat{r}_\rho$ . As illustrated in Figure 9, the defect zones were well discriminated, we can thus binarize the point cloud by a threshold. Since the histogram of the distance was unimodal, we used a unimodal thresholding method proposed by Rosin in [17]. By this method, a straight line is drawn from the maximum to the first empty bin of the histogram, to form

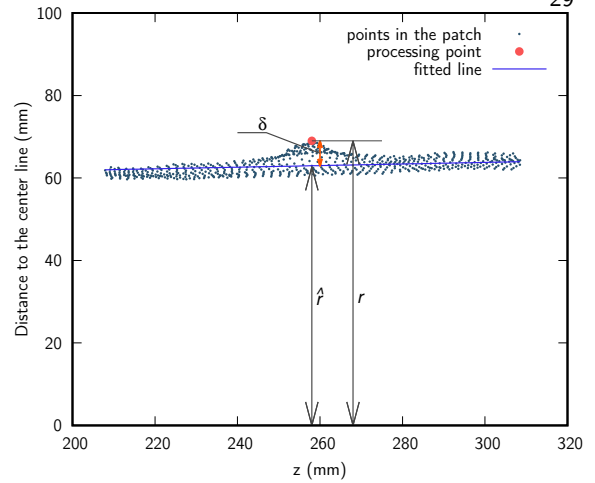


Figure 8. Computation of the reference distance to the centerline for a point

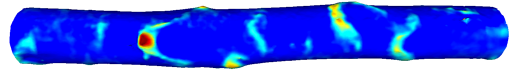


Figure 9. The distance map of  $\delta$ , in green are the regions with a small value, in red are the regions with a greater value

a triangular-like shape. The threshold  $T$  is selected at the point of the histogram that maximizes the perpendicular distance between the histogram and the straight line (see Figure 10). The point cloud is then binarized by this threshold: with  $\delta(\rho)$  is the difference between distance to the centerline and reference distance to the centerline of the point  $\rho$ , the function  $g(\rho)$  determines whether  $\rho$  is a defect point or not.

$$g(\rho) = \begin{cases} 1 & \text{if } \delta(\rho) > T \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

## III. EXPERIMENTAL RESULTS

In order to evaluate the proposed method, we carried out experiments involving the point clouds describing 10

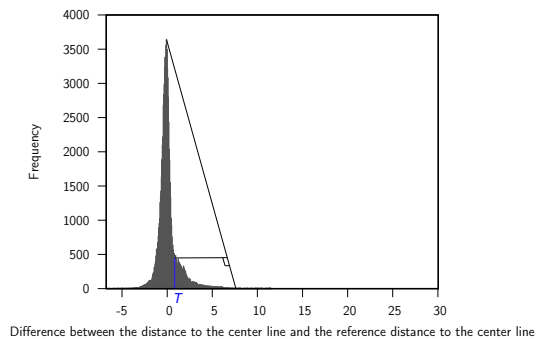


Figure 10. Histogram of the difference between the distance to the centerline and the reference distance to the centerline. Due to the different scale between the two axis, the two perpendicular lines are not shown correctly.

logs from 7 different tree species. The principal objective was to test the robustness of the method with different tree species, flexuosities, cross-sectional shapes, bark roughness, and defect sizes. The defects on each log were manually segmented to create the ground truth. The proposed method was also compared with a cylinder fitting based method which is similar to the approach used in [9], [12]. We fitted the successive cylinders of 14 cm length, then we computed the distance from the points to the surfaces of these cylinders. The same thresholding method [17] was used to binarize the log point cloud.

Visually, the segmented defects corresponded well to the ground truth, even with flexuosity (see Figure 11 (a)), or with a rough bark (see Figure 11 (b)). However, we observed that in the latter case, some small prominent parts of the bark were miss-classified as defects. The same observation was made with a log which did not have large defects: the most prominent parts of the bark were miss-classified as in the example of the Wild service tree log shown in Figure 11 (c). Note that most of these miss-classified defects can be eliminated by a post-process by filtering out the region with a significant small area or by a priori knowledge about defect shapes. For example, most of the defects have the Chinese mustache shape, ellipse or circle, the miss-classified defects often have an elongated shape.

In order to assess the segmentation performance in a quantitative manner, we use the classic metrics: precision, recall and F-measure (F1). As shown in the table I, with an overall F-measure 0.710. While the method achieved a pretty stable recall ranging from 0.604 to 0.833, the precision of some logs was lower for the above-mentioned reasons (Red oak 2, Wild service tree).

The proposed method outperforms the cylindrical-based method on all tested logs. Figure 12 shows a comparison between the proposed method and the cylindrical-based method on Wild cherry logs. On the first line, we see that the defect detected by the proposed method is more precise than the cylindrical one. While both methods could segment the scar part of the defect, the mustache parts were only segmented by the proposed method. This information is important to evaluate the size of internal defect [12], [10]. We also observed that the proposed method could detect smaller defects with very weak relief as shown in the second line of the Figure 12.

The results also showed that, on the logs which had a complex cross-section shapes as the Fir 1 log and the Red oak 2 log shown in the Figure 13, while the proposed method retained a good performance (F-measures respectively equal 0.757 and 0.56), the cylindrical-based method had a bad performance (F-measures respectively equal 0.238 and 0.104).

In the experiment with the roughest log available in the database, the Elm log, we obtained a F-measure 0.721 which is slightly greater than the overall performance and much superior than the cylindrical-based one. This result proved that the bark roughness had little effect on the proposed method. On some defect caused by a living branch, the cylindrical-based method may perform

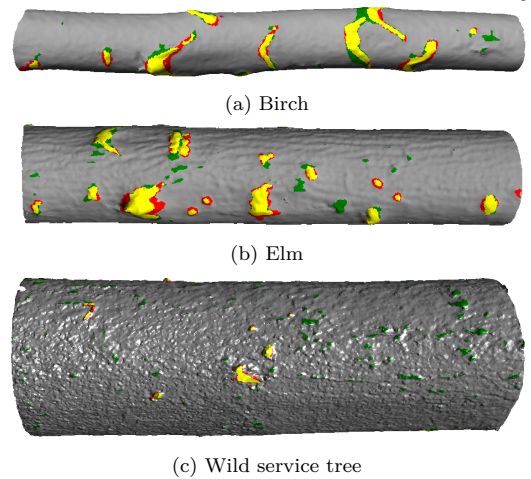


Figure 11. Segmentation results of the proposed method for some logs: the detected zones, overlaying the ground truth are in yellow, non overlaying the ground truth are in green, and the zones in ground truth but non-detected are in red.

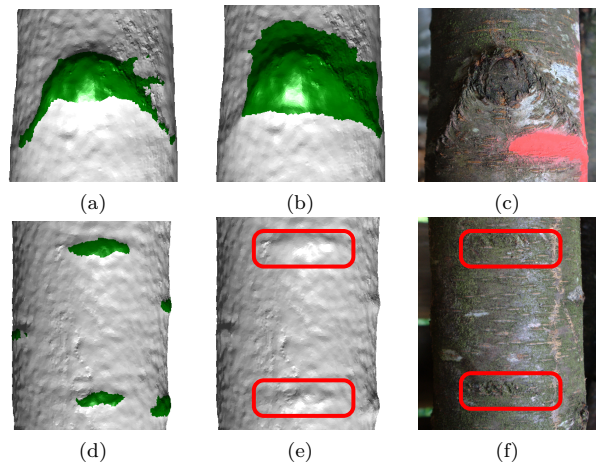


Figure 12. Comparison between the results of the proposed method (first column) and the cylindrical-based method (second column) on a large and a small size defects. Photographs of the defects are presented in the third column.

better than the proposed method in the segmentation of the branch part. Nevertheless, the Chinese mustaches were only segmented by the proposed method (see Figure 14). However, in this work, we did not expect to have the best performance on this defect type, which can be segmented by another way.

The results showed that our method can precisely segment the defects and seems to be robust to different tree species and geometrical features and must be preferred to the cylindrical-based method. Note that these results can be reproduced and compared on an online demonstration [18].

#### IV. CONCLUSION

This paper proposed a novel approach for segmenting the defects on the surface of tree logs from Terrestrial

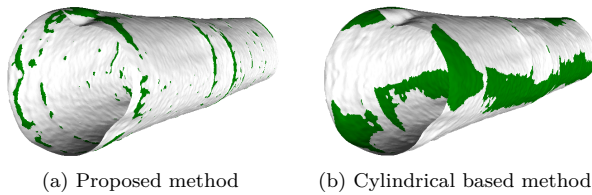


Figure 13. Segmentation performance on the less cylindrical log Red oak 2. Green zones are segmented defects.

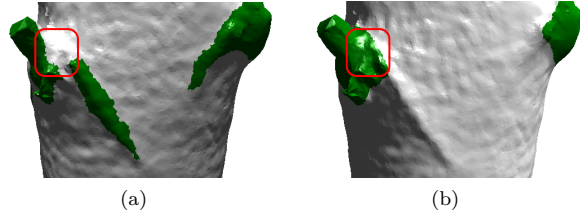


Figure 14. The living branch is partially segmented by the proposed method (a). While the cylindrical-based method can fully segment it, the mustache is not segmented (b).

Table I. RESULT OF THE SEGMENTATION ALGORITHM

Log	Proposed method			Cyl.-based method[12]		
	Prec <sup>1</sup>	Recall	F1	Prec	Recall	F1
Fir 1	<b>0.747</b>	0.769	<b>0.757</b>	0.137	<b>0.937</b>	0.238
Fir 2	<b>0.673</b>	<b>0.775</b>	<b>0.719</b>	0.353	0.452	0.395
Wild cherry 1	<b>0.696</b>	<b>0.765</b>	<b>0.728</b>	0.683	0.512	0.584
Wild cherry 2	<b>0.846</b>	0.711	<b>0.771</b>	0.661	<b>0.822</b>	0.732
Red oak 1	<b>0.749</b>	<b>0.742</b>	<b>0.744</b>	0.479	0.444	0.459
Red oak 2	<b>0.428</b>	<b>0.833</b>	<b>0.564</b>	0.061	0.400	0.104
Beech	<b>0.670</b>	<b>0.604</b>	<b>0.634</b>	0.360	0.289	0.320
Birch	<b>0.733</b>	<b>0.756</b>	<b>0.744</b>	0.607	0.421	0.496
Elm	<b>0.694</b>	<b>0.755</b>	<b>0.721</b>	0.494	0.309	0.378
Wst <sup>2</sup>	<b>0.247</b>	<b>0.741</b>	<b>0.370</b>	0.057	0.463	0.100
Overall	<b>0.685</b>	<b>0.740</b>	<b>0.710</b>	0.289	0.563	0.380

<sup>1</sup> Precision

<sup>2</sup> Wild service tree

Laser Scanning data. Unlike the previous methods based on primitive shape fitting, the proposed method is based on a precise computation of the log centerline, and for each point, a patch containing its neighbors is used to estimate the reference distance to the centerline. The experiments showed that the method could precisely segment defects with complex shape like the Chinese mustaches and small defects. Moreover, by comparing with a method based on cylindrical fitting, the results were more efficient for different tree species and for logs deviating from the cylindrical model. However, in the case of logs with very prominent defect like a living branch or without significant defect, the proposed method did not have a good performance. Processing of these particular cases is an ongoing work. The proposed method can be used in processing chain aiming the characterization of internal defects from outside for assessing the quality of logs or of standing trees.

**Acknowledgment:** the authors thank of the **DGtal** library [19] allowing to generate the 3D illustrations.

This work was supported by the French National Research Agency through the Laboratory of Excellence ARBRE (ANR-12- LABXARBRE-01) and by the Lorraine French Region.

## REFERENCES

- [1] B. Kerautret, A. Krähenbühl, I. Debled-Rennesson, and J.-O. Lachaud, "3d geometric analysis of tubular objects based on surface normal accumulation," in *Image Analysis and Processing—ICIAIP 2015*. Springer, 2015, pp. 319–331.
- [2] U. Bauer and K. Polthier, "Generating parametric models of tubes from laser scans," *Computer-Aided Design*, vol. 41, no. 10, pp. 719–729, Oct. 2009.
- [3] C. Kirbas and F. Quek, "A review of vessel extraction techniques and algorithms," *CSUR*, vol. 36, no. 2, pp. 81–121, 2004.
- [4] D. Lesage, E. D. Angelini, I. Bloch, and G. Funka-Lea, "A review of 3d vessel lumen segmentation techniques: Models, features and extraction schemes," *Medical Image Analysis*, vol. 13, no. 6, pp. 819–845, Dec. 2009.
- [5] O. Tankyevych, H. Talbot, N. Passat, M. Musacchio, and M. Lagneau, "Angiographic image analysis," in *Medical Image Processing*. Springer, 2011, pp. 115–144.
- [6] A. Krähenbühl, B. Kerautret, I. Debled-Rennesson, F. Mothe, and F. Longuetaud, "Knot Segmentation in 3d CT Images of Wet Wood," *Pattern Recognition*, 2014.
- [7] F. Longuetaud, F. Mothe, B. Kerautret, A. Krähenbü, L. Hory, J. M. Leban, and I. Debled-Rennesson, "Automatic knot detection and measurements from X-ray CT images of wood: A review and validation of an improved algorithm on softwood samples," *COMPAG*, vol. 85, pp. 77–89, 2012.
- [8] R. E. Thomas, "Modeling the relationships among internal defect features and external appalachian hardwood log defect indicators," 2009.
- [9] S. M. Stängle, F. Brüchert, U. Kretschmer, H. Spiecker, and U. H. Sauter, "Clear wood content in standing trees predicted from branch scar measurements with terrestrial lidar and verified with x-ray computed tomography 1," *Canadian Journal of Forest Research*, vol. 44, no. 2, pp. 145–153, 2013.
- [10] V. Racko, "Verify the accuracy of estimation the model between dimensional characteristics of branch scar and the location of the knot in the beech trunk," *Annals of Warsaw Univ. of Life Sciences-SGGW. Forestry & Wood Technology*, vol. 84, 2013.
- [11] C. Schütt, T. Aschoff, D. Winterhalder, M. Thies, U. Kretschmer, and H. Spiecker, "Approaches for recognition of wood quality of standing trees based on terrestrial laserscanner data," *ISPRS*, vol. 36, pp. 179–182, 2004.
- [12] U. Kretschmer, N. Kirchner, C. Morhart, and H. Spiecker, "A new approach to assessing tree stem quality characteristics using terrestrial laser scans," *Silva Fenn*, vol. 47, p. 14, 2013.
- [13] L. Thomas, C. A. Shaffer, L. Mili, and E. Thomas, "Automated detection of severe surface defects on barked hardwood logs," *Forest products journal*, vol. 57, no. 4, p. 50, 2007.
- [14] L. Thomas and R. E. Thomas, "A graphical automated detection system to locate hardwood log surface defects using high-resolution three-dimensional laser scan data," in *17th Central Hardwood Forest Conference*, vol. 78, 2010, p. 92.
- [15] L. Thomas and L. Mili, "A robust gm-estimator for the automated detection of external defects on barked hardwood logs and stems," *Signal Processing, IEEE Transactions on*, vol. 55, no. 7, pp. 3568–3576, 2007.
- [16] B. Gough, *GNU scientific library reference manual*. Network Theory Ltd., 2009.
- [17] P. L. Rosin, "Unimodal thresholding," *Pattern recognition*, vol. 34, no. 11, pp. 2083–2096, 2001.
- [18] "Online demonstration," [http://ipol-geometry.loria.fr/~kerautre/ipol\\_demo/TrunkDefaultMeasure](http://ipol-geometry.loria.fr/~kerautre/ipol_demo/TrunkDefaultMeasure).
- [19] "DGtal: Digital Geometry tools and algorithms library," <http://dgtal.org>.

## **2.6 Article 2 : Algorithms and Implementation for Segmenting Tree Log Surface Defects**

Publié dans les actes de l'International Workshop on Reproducible Research in Pattern Recognition, Springer, pp 150–166

# Algorithms and Implementation for Segmenting Tree Log Surface Defects

Van-Tho Nguyen<sup>1</sup>, Bertrand Kerautret<sup>2</sup>, Isabelle Debled-Rennesson<sup>2</sup>, Francis Colin<sup>1</sup>, Alexandre Piboule<sup>3</sup>, and Thiéry Constant<sup>1</sup>

<sup>1</sup> LERFOB, AgroParisTech, INRA, 54000 NANCY France

<sup>2</sup> LORIA, UMR CNRS 7503, Université de Lorraine, 54506 Vandœuvre-lès-Nancy, France

<sup>3</sup> ONF, RDI, 5 Rue Girardet, 54000 Nancy, France.

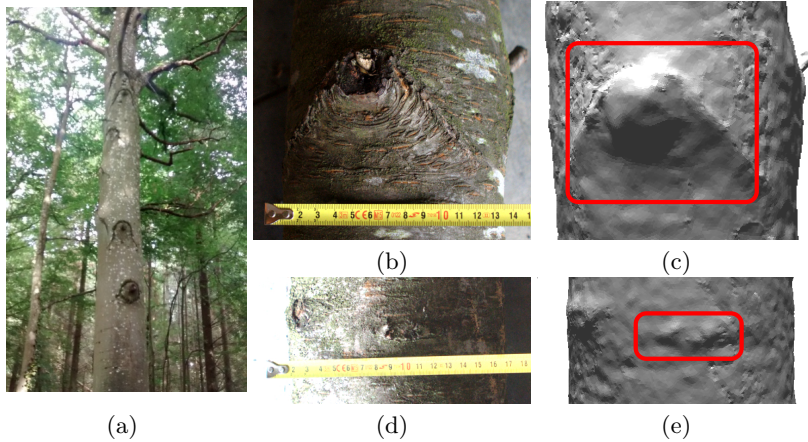
**Abstract.** This paper focuses on the algorithms and implementation details of a published segmentation method defined to identify the defects of tree log surface. Such a method overcomes the difficulty of the high variability of the tree log surface and allows to segment the defects from the tree bark. All the algorithms used in this method are described in link to their source code which guarantees a full reproducible method associated to an online demonstration.

## 1 Overview of the segmentation of defects on log surface

In the computer imagery domain, the tubular objects are present in various applications of which segmentation and analysis are of most importance like in medicine (with the blood vessel segmentation [9,12]), biology (with the wood knot detections [11]) or industry [4,7]. In this work, we focus on the problem of wood quality estimation in relation to the presence of internal wood knots. Since the latter have a structural link with defects on the surface of the trunk, the non destructive Terrestrial Laser Scanning (TLS) can be used to obtain an estimation of the wood quality. As described in previous work [14], different approaches were proposed following this strategy with for instance the fitting of primitive such as cylinder [17,18,10] or circle [20,19]. However, the existing approaches are not fully satisfactory in regards to the specificity of some species [20,21] or for the automation of the method [17,18,10]. Another limitation is the implementation details which are often missing and limit the reproduction of the methods.

A generic method to segment automatically defects on tree log surface that is robust to various tree species is still missing. In [14], we presented a novel approach to segment defects on tree logs from TLS data that is robust to different tested tree species in regard to bark roughness, longitudinal flexuosity or cross-sectional ovality. However, being limited on the number of pages, we were not able to provide enough details about algorithms, implementations and also how to reproduce the best results. As a complement of [14], this paper aims to provide the implementation details of the main algorithms and to discuss the choice of

parameter values and the reader can also access to the source code and other resources to reproduce the published results.



**Fig. 1.** (a) Defects on a standing tree in forest. (b,c) A Chinese mustache and its corresponding mesh. (d,e) A small defect in ellipse shape and its corresponding mesh.

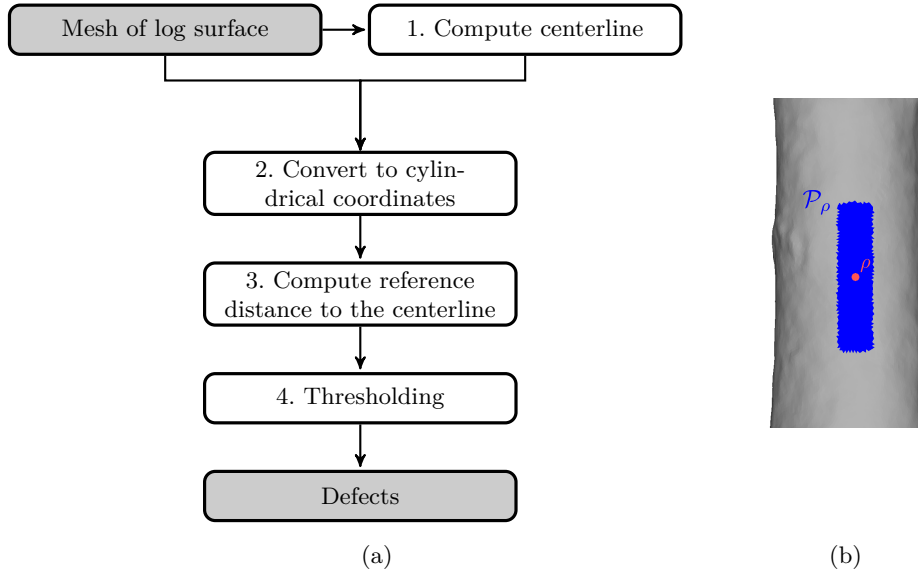
Our approach to segment tree log surface defects consists of four main steps (see Fig. 2 (a)). In the first step, we compute the centerline of the tree log by the accumulation of surface normal vectors. In the second step, based on this centerline, we can compute the distance to the centerline and then convert the point cloud from the cartesian coordinate system to cylindrical coordinate system. The third step concerns the computation of the reference distance to the centerline using a patch of neighbouring points. Finally, in the fourth step we compute the difference between the reference and the real distances to the centerline before applying an automatic threshold to binarize the point cloud according to the statistical distribution of the values.

The following section describes the details and implementations of our algorithms. The section 3 describes the steps to reproduce the results. We discuss how to choose the best values of the most important parameters and show some limitations of our method in the section 4.

## 2 Algorithms and implementations

In this section, we describe in detail our algorithms and its implementations. The program is written in *C++* using libraries DGtal [1] for the normal accumulation and the visualization, PCL [16] for the kd-tree, GNU Scientific Library (GSL) [6] for the smoothing and interpolation of the centerline by cubic Spline. Fig. 3 shows main classes and methods that will be described in detail in this section. By convenience, our source files have the same name as class names. For example





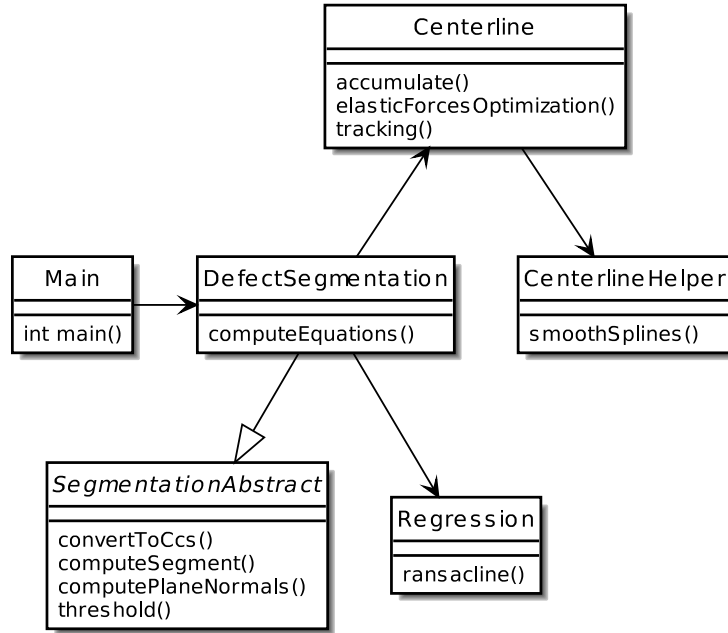
**Fig. 2.** (a) Overview of the algorithm. (b) Illustration of a patch  $\mathcal{P}_\rho$  in blue which is used to estimate the reference distance to the centerline for the orange point.

the class `Centerline` is implemented in the file `Centerline.h` and `Centerline.cpp`.

## 2.1 Compute centerline

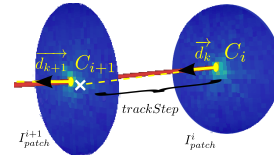
In the proposed method, the centerline plays an important role because the key variable is its distance to the input points. Thus, the centerline must be precisely assessed. Our method to compute the centerline is based on the surface normal accumulation proposed in [7] with modifications optimized to tree log data. The implementation of this method is simple (Fig. 4 (a)) and consists of four steps: (i) accumulation of surface normal vector, (ii) tracking of the centerline, (iii) optimization by elastic forces (see Algorithm 1) and (iv) BSpline interpolation.

**(i) Accumulation of surface normal vectors.** This step consists in computation of a volumetric image of which voxel value stores score of the accumulation (Fig. 4 (b)) which is computed as follows. Starting from a mesh face  $f_i$  with its normal vector  $\vec{n}_i$  and applying a normal oriented directional scan along a distance  $d_{acc}$  (oriented toward the object interior). During this scan, we increase by one all voxels of the digitized space which are crossed by the 3D line defined from the face center and from the direction  $\vec{n}_i$  (see Fig. 4 (a,b)) and Algorithm 1 of [7] for more details).



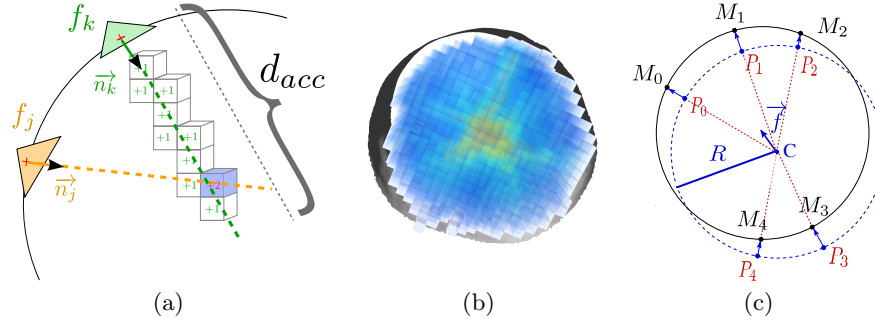
**Fig. 3.** Diagram class of our implementation.

(ii) **Tracking the centerline.** Based on the result of previous algorithm, the centerline curve is obtained by a simple tracking process by looking at the local maxima accumulation points  $C_i$  of 2D patch image resulting from the projection in the  $\vec{d}_k$  direction (see following figure). This direction  $\vec{d}_k$  is computed from all the points which are at the origin of the accumulation value (see Algorithm 2 of [7] for more details).



(iii) **Elastic force optimization.** Due to the digital space and the mesh quality, the normal vectors may not be perfectly convergent and some irregularities on the resulting centerline can appear. As suggested in [7], we improve this step by using an optimization algorithm Algorithm 1. The main idea is to minimize the error  $E_s(C)$  defined as the sum of the squared difference between the mean radius  $\bar{R}$  of the log and the norm of the projection vector formed by log center  $C_i$  and its associated input mesh points  $M_i$  onto the plane perpendicular to the direction vector  $\vec{d}_i$  (see Fig. 4 (c)).

(iv) **Spline Interpolation.** The purpose of the interpolation is to obtain a smoother centerline. We firstly choose eight points on the centerline based on



**Fig. 4.** Illustrations of the main ideas of the centerline method based on surface normal accumulation: (a) scan from mesh faces along a distance  $d_{acc}$ , (b) resulting accumulation score. (c) illustrates the step of elastic force optimization.

the curvature. The points with a too strong curvature change are eliminated by testing the angle  $\phi$  defined by the considered point, and its two immediate neighbors on both sides along the centerline. If  $\phi$  is smaller than  $3\pi/4$  then the point is eliminated. Finally, a cubic spline interpolation is applied to these chosen points. Our implementation of this step is based on cubic spline which is implemented in GNU scientific library [6]. Fig. 5 (a) shows the final result of a computed centerline.

## 2.2 Convert to cylindrical coordinates

The purpose of this conversion is to simplify the search of neighbors. Firstly, we divide the point cloud into slices by the centerline segment computed in the previous section. Each point in the point cloud belongs to only one slice (Algorithm 2). As shown in [14], we define a local coordinate system  $O_i x_i y_i z_i$  for each slice (see Algorithm 4) where  $\{\vec{u}_i, \vec{v}_i, \vec{w}_i\}$  are basis vectors with the origin  $O_i$  (i.e the first point on the segment of the centerline corresponding to the slice).  $\vec{w}_i$  is defined as the segment  $C_i C_{i+1}$ . The axis  $O_0 y_0$  can be arbitrary and we choose  $O_0 y_0 = Oy$ . With  $i > 0$ ,  $\vec{v}_i$  is computed as follows. Let  $\vec{n}_i$  be the normal vector of the plane formed by  $\vec{w}_i$  and  $\vec{v}_{i-1}$ :  $\vec{n}_i = \vec{w}_i \times \vec{v}_{i-1}$ ;  $i \in [1, m-1]$ .  $\vec{v}_i$  is the cross product of  $\vec{w}_i$  and  $\vec{n}_i$ . And finally,  $\vec{u}_i$  is computed by the cross product of  $\vec{v}_i$  and  $\vec{w}_i$ . Fig. 5 (b) shows the computation of coordinate system for a point from its Cartesian coordinates.

## 2.3 Compute reference distance to the centerline

To compute the reference distance to the centerline of the point  $\rho$ , we need to query a narrow patch  $\mathcal{P}_\rho$  containing the neighbors of  $\rho$  (see Fig. 2 (b)). The patch size is defined by two parameters  $\varphi$  and  $\tau$  with  $\varphi = l/\bar{r}$  where  $\tau$  (resp.  $l$ ) is the length (resp. arc length) of the patch and  $\bar{r}$  is the mean of the distance to the centerline of all points. More formally the patch  $\mathcal{P}_\rho$  can be defined as:

---

**Algorithm 1: elasticForcesOptimisation:** Optimize the centerline by elastic forces
 

---

```

Data: rawCenterline // Raw centerline, output of Algorithm 2 of [7]
Data: mesh // Input mesh to recover faces associated to centerline points
Output : optimizedCenterline
Variable: epsilon //Error
1 //store associated faces in the 2D patches
2 mapFaces = emptyMap()
3 sumradii=0
4 foreach point in rawCenterline do
5   faces = get2DpatchImage(point, mesh)
6   mapFaces[point] = faces
7   foreach face in faces do
8     centerPoint = face.center()
9     vectorNormal = face.getVectorNormal()
10    vectorRadial = centerPoint - point
11    sumradii = vectorRadial.norm()

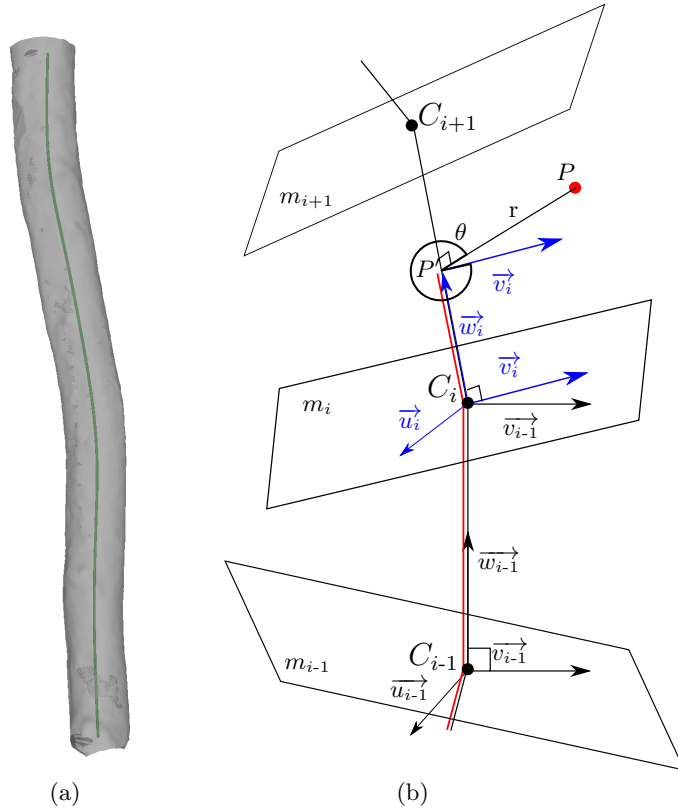
12 meanRadii = sumradii / mesh.nbFaces()
13 DeltaError = infinity
14 previousTotalError
15 first=true
16 while DeltaError > epsilon do
17   totalError = 0.0
18   for i = 0 to rawCenterline.size() - 1 do
19     point = optimizedCenterline[i]
20     faces = mapFaces[point]
21     sumForce = {0,0,0}
22     count=0
23     foreach face in faces do
24       centerPoint = face.center()
25       vectorNormal = face.getVectorNormal()
26       radialVector = centerPoint - point
27       alpha = anglebetween(vectorNormal, radialVector)
28       if alpha >  $\frac{\pi}{6}$  then
29         | continue
30       forceMagnitude = radialVector.norm() - meanRadii
31       force = radialVector.normalized()*forceMagnitude
32       totalError += forceMagnitude*forceMagnitude
33       sumForce += force
34     count++
35     //project of sumForces to normal vector
36     direction = dirImage[rawCenterline[i]].normalized()
37     sumForcesDir = direction.dot(sumForces)/vectDir.norm()/vectDir.norm()*vectDir;
38     radialForces = sumForces - sumForcesDir;
39     if count > 0 then
40       | optimizedCenterline[i] += radialForces/count

41 if first then
42   | DeltaError = totalError first = false
43 else
44   | DeltaError = abs(totalError - previousTotalError)
45   | previousTotalError = totalError - previousTotalError
46 return optimizedCenterline
47 .

```

---

$$\mathcal{P}_\rho = \{\rho_j \mid |\theta_{\rho_j} - \theta_\rho| \leq \frac{\varphi}{2}, |z_{\rho_j} - z_\rho| \leq \frac{\tau}{2}\} \quad (1)$$



**Fig. 5.** (a) Centerline of a Beech log. (b) Computation of the coordinate system  $\rho(r, \theta, z)$  from the point  $P$  with  $z$  is the red line.

---

**Algorithm 2: computePlanes:** Compute normal vector of planes that divide the log into slices

---

**Data:** Centerline  $\mathcal{C} = \{C_0, C_1, \dots, C_m\}$   
**Result:** List of normal vector  $m$

```

1  $m = \text{emptySet}()$ 
2  $m[0] = \frac{\overrightarrow{C_0 C_1}}{\|\overrightarrow{C_0 C_1}\|}$ 
3 for  $i = 1$  to  $m - 1$  do
4    $m[i] = \frac{\overrightarrow{C_{i-1} C_i} + \overrightarrow{C_i C_{i+1}}}{\|\overrightarrow{C_{i-1} C_i} + \overrightarrow{C_i C_{i+1}}\|}$ 
5 end
6 return  $m$ 

```

---

Because the number of points may be large, we use a kd-tree to speedup the query. We use the kd-tree implemented in the *Flann* library [3,13] which is included in the *PCL* library [16]. This library provides a range base query, so

---

**Algorithm 3: getSegmentId:** Compute the corresponding segment of a point

---

**Data:** Point  $P$   
**Data:** Centerline  $\mathcal{C} = \{C_0, C_1, \dots, C_m\}$   
**Result:** segmentId

```

1 lastSign ← 1
2 for i = 1 to m do
3   sign =  $\overrightarrow{C_i P} \cdot ns[i]$ ;
4   if sign * lastSign ≤ 0 then
5     return i - 1;
6   end
7   lastSign = sign;
8 end
9 return m - 1
```

---



---

**Algorithm 4: convertToCcs:** Convert the point cloud to cylindrical coordinate system

---

**Data:** Point cloud  $\mathcal{P}_d = \{P_0, P_1, \dots, P_n\}$   
**Data:** Centerline  $\mathcal{C} = \{C_0, C_1, \dots, C_m\}$   
**Data:** Local coordinate systems  $\mathcal{L} = \{O_0 x_0 y_0 z_0, O_1 x_1 y_1 z_1, \dots, O_{m-1} x_{m-1} y_{m-1} z_{m-1}\}$   
**Result:** Point cloud in cylindrical coordinate system

```

1 for i = 1 to n do
2   s = getSegmentId(Pi) //Algorithm 2
3    $\vec{d} = \frac{\overrightarrow{C_s C_{s+1}}}{\|\overrightarrow{C_s C_{s+1}}\|}$  //projection of Pi onto segment s of the centerline
4    $P' = \vec{d} \cdot \overrightarrow{C_s P_i}$ ;
5    $r_i = \|\overrightarrow{P' P_i}\|$ 
6   if s > 0 then
7      $z_i = \sum_{j=1}^s \|\overrightarrow{C_{j-1} C_j}\| + \overrightarrow{C_j P_i} \cdot \vec{d}$ 
8   else
9      $z_i = \overrightarrow{C_j P_i} \cdot \vec{d}$ 
10  end
11   $\theta = \frac{\arccos \overrightarrow{P' P_i} \cdot \vec{v}_s}{\|\overrightarrow{P' P_i}\|}$ 
12  //correction of angle
13   $\vec{t} = \vec{v}_s \times \vec{d}$ 
14  if  $\vec{t} \cdot \overrightarrow{P' P_i} < 0$  then
15     $\theta = 2\pi - \theta$ ;
16  end
17 end
```

---

that at the point  $\rho$ , we query all the neighbors located inside the sphere of center  $\rho$  and  $radius = \frac{\tau}{2}$ :

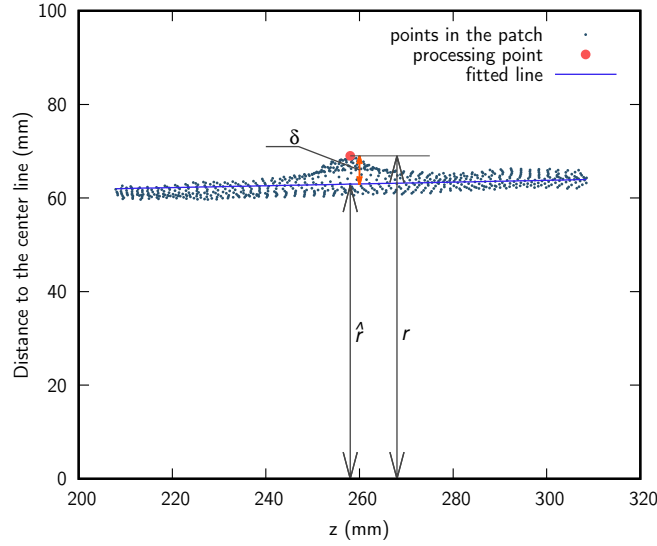
$$\mathcal{Q}_\rho = \{\rho_j \mid |z_{\rho_j} - z_\rho| \leq \frac{\tau}{2}\} \quad (2)$$

The return must be refined to get the patch:

$$\mathcal{P}_\rho = \{\rho_j \in \mathcal{Q} \mid |\theta_{\rho_j} - \theta_\rho| \leq \frac{\varphi}{2}\} \quad (3)$$

Note that the query of patch can be improved by using a kd-tree for the  $\theta$  and  $z$  coordinates in the cylindrical coordinate system with a regard of the circular problem of  $\theta$ .

To compute the reference distance to the centerline (denoted  $\hat{r}$ ) of the point  $i$ , we consider the profile of  $r$  by  $z$  and then compute a RANSAC [5] based linear regression (i.e  $\hat{r} = az + b$ , see Fig. 6).



**Fig. 6.** Computation of the reference distance to the centerline for a given point.

## 2.4 Thresholding

After computing the reference distance to the centerline of points in the point cloud, we use the unimodal thresholding method proposed by Rosin in [15] to binarize the point cloud. Firstly, at each point  $\rho$ , we compute the difference between the distance to the centerline  $r$  and the reference distance to the centerline  $\hat{r}$  for all points ( $\delta_\rho = r_\rho - \hat{r}_\rho$ ).

Then, we compute the histogram of  $\delta_\rho$  of all points in the cloud. The algorithm firstly finds the bin with maxima frequency ( $B_1$ ) and then finds the first null bin ( $B_2$ ) at the right size of the histogram. Finally, the algorithm loops over the bins situated between  $B_1$  and  $B_2$  and chooses the bin that maximizes the distance to the line  $B_1B_2$  (Fig. 7). The Algorithm 5 shows the details of the implementation.

---

**Algorithm 5: threshold:** Compute threshold by the Rosin method
 

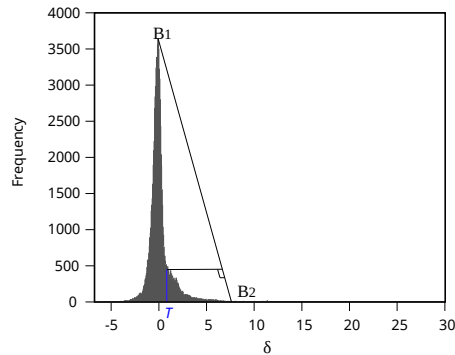
---

```

Data:  $\mathcal{D} = \{\delta_0, \delta_1, \dots, \delta_n\}$ 
Data:  $binSize$ 
Result: threshold  $T$ 
1 nbBins =  $\frac{max(\mathcal{D}) - min(\mathcal{D})}{binSize}$ 
2 histogram = array[nbBins]
3 for  $i = 1$  to  $n$  do
4   binId =  $\frac{\rho_i - min(\mathcal{D})}{binSize}$ 
5   histogram[binId]++
6 end
7 maxFrequency = max(histogram)
8 maxFrequencyIndex = indexOf(maxFrequency)
9 nullFrequencyIndex = nbBins - 1
10 for  $i = maxFrequencyIndex$  to  $nbBins$  do
11   if histogram[i] == 0 then
12     nullFrequencyIndex = i
13     break
14   end
15 end
16 bestDist = 0
17 bestDistIndex = maxFrequencyIndex
18 for  $i = maxFrequencyIndex$  to  $nullFrequencyIndex$  do
19   AB = line({maxFrequencyIndex, maxFrequency}, {nullFrequencyIndex, 0})
20   if distance(i, histogram[i], AB) > bestDist then
21     bestDist = distance(i, histogram[i], AB)
22     bestDistIndex = i
23   end
24 end
25 return  $T = bestDistIndex * binSize + min(\mathcal{D})$ 

```

---



**Fig. 7.** Computation of the threshold  $T$ . Due to the different scales on the two axis, the two perpendicular lines are not shown correctly.

### 3 Reproducing the results

The source code used to generate the present results is given at the following *GitHub* repository:

<https://github.com/vanthonguyen/treelogdefectsegmentation>

and an online demonstration is available at the following url:



[http://ipol-geometry.loria.fr/~kerautre/ipol\\_demo/TDD\\_IPOLDemo/](http://ipol-geometry.loria.fr/~kerautre/ipol_demo/TDD_IPOLDemo/)

The program was compiled and tested on GNU linux operating system and on MacOS. The recommended compiler is Gcc 4.7 or later. It could be compiled and run in windows operating system, however some additional software should be installed like mingw or MS C++ compiler. The program requires these libraries to be installed: DGtal version 0.9, PCL version > 1.7, Eigen3, GNU GSL and Boost. Once the program is compiled, the command to run is:

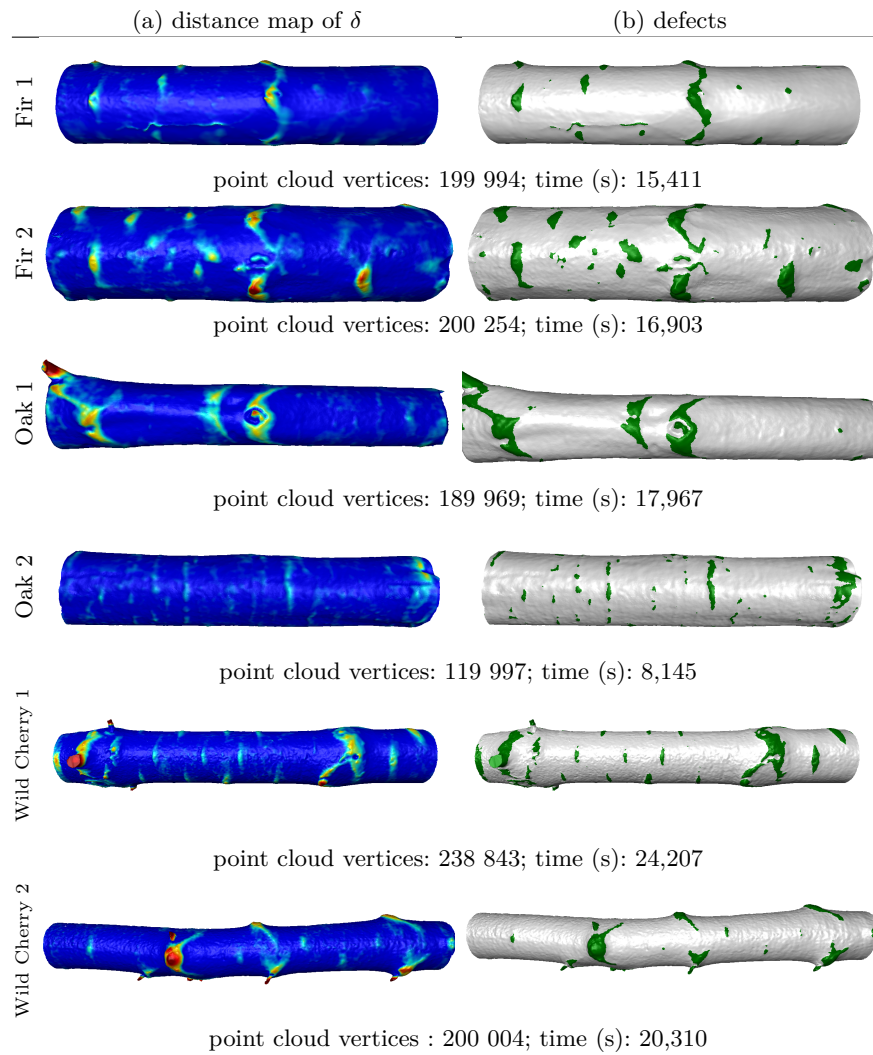
```
./segmentation -i mesh.off --voxelSize 5 --accRadius 100 --trackStep 20 \
--patchWidth 25 --patchHeight 100 --binWidth 5 --invertNormal true -o prefix
```

The program accepts the following parameters:

- **voxelSize**: the voxel size which is related to the resolution of the point cloud. In our experiments, the density of the point cloud is about 25 points per  $\text{cm}^2$ , we have chosen the voxel size as the squared root of the density which is 5.
- **accRadius**: the radius of the normal accumulation which should be greater than the real radius of the log.
- **trackStep**: the distance between two points in the tracking of the centerline.
- **patchWidth**, **patchHeight**: the width and height of the patch in the search of neighbors. The choice of the patch size must guarantee that the height is several time greater than the width and that the width is enough large to avoid an empty patch. In our experiments, we fixed the width equals to 25 mm and the height equals to 100 mm.
- **binWidth**: the width of histogram bins used to compute the threshold by the Rosin method.
- **invertNormal**: used when the direction of normal vectors is outside of the object.
- **output**: the prefix of output files. The program writes the output on some files: (1) output mesh with highlighted defects in green (*prefix-defect.off*), (2) the distance map of  $\delta$  (*prefix-error.off*), (3) the face ids of defects (*prefix-def-faces.id*) which can be used with the tool *colorizeMesh* (included in source code), (4) the point ids of defects (*prefix-defect.id*) which can be used to compare with the ground truth. The format of these two last files are simply a list of integers separated by newlines.

The results obtained on different tree species including ground truth error measures were already presented in previous work [14]. We focus in this part on the reproduction of the previous results and in the experimental stability measure of the different parameters.

**Reproducing results on various species.** From the previous command line, the method can be applied directly to different species without the need to change the default parameters. The Fig. 8 presents some results generated from the example directory of the *GitHub* repository. All the results were generated with the same default parameters: **voxelSize**= 5mm, **accRadius** is chosen by 1.5 time the maximum radius of the log (by default set to 200), the **patchWidth**=25 mm,



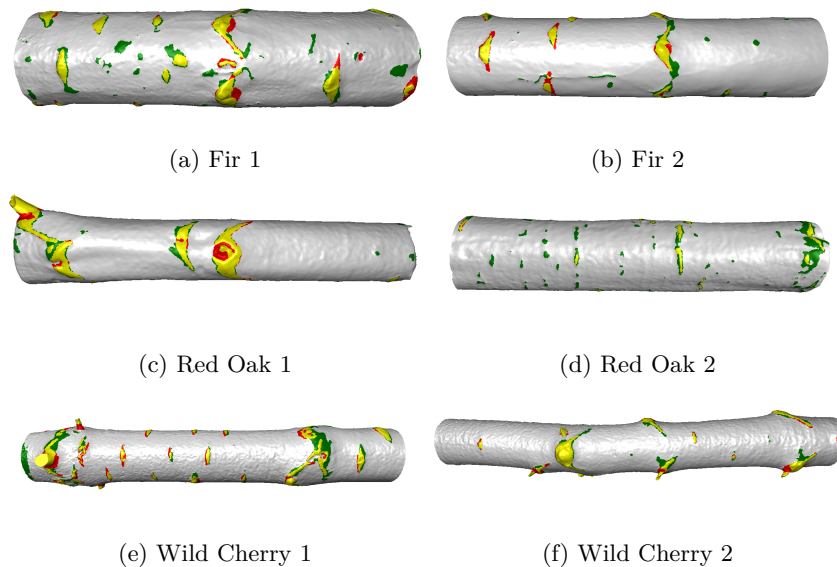
**Fig. 8.** Visualization of the distance map of  $\delta$  (column (a)) and defects (column (b)) obtained on various species with the same default parameters (`voxelSize=5`, `accRadius=200`, `trackStep=20`, `patchWidth=25`, `patchHeight=100`, `binWidth=0.01`). The execution time were obtained on a *MacOS* 2,5 GHz *Intel Core i7*.

`patchHeight=100` mm. Note that the similar results are available on the demo site [2] with other examples.

**Reproducing ground-truth comparisons.** The figure Fig. 9 illustrates the comparisons of the previous results with the ground truth constructed by INRA experts built from logs of different tree species and diameters, with a one meter length. Such comparisons can also be reproduced from the following command line:

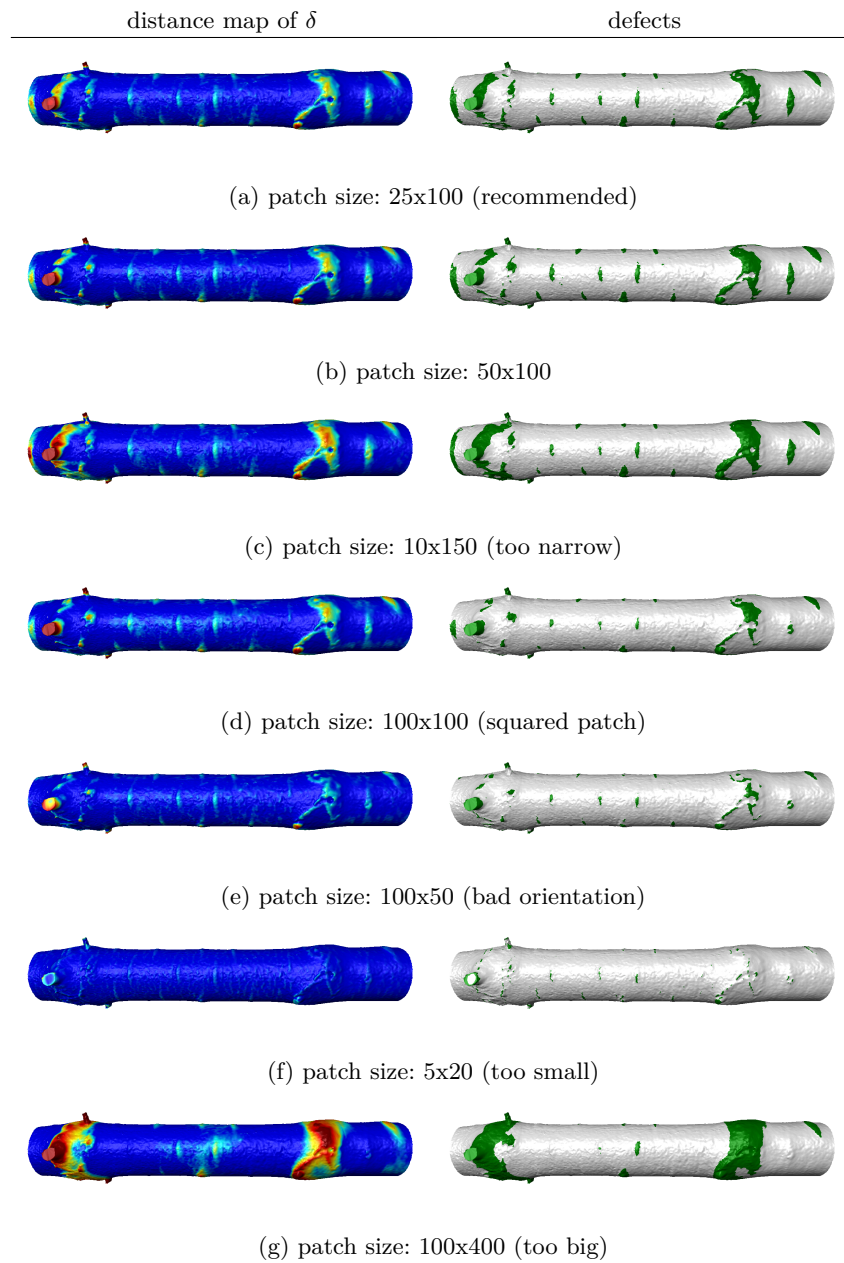
```
colorizeMesh -i examples/WildCherry1.off -r examples/WildCherry1-grountruth.id
-t res_WildCherry-def-faces.id -o compareWildCherry.off
```

To ensure reproducibility the ground-truth files are also given in the *GitHub* repository.



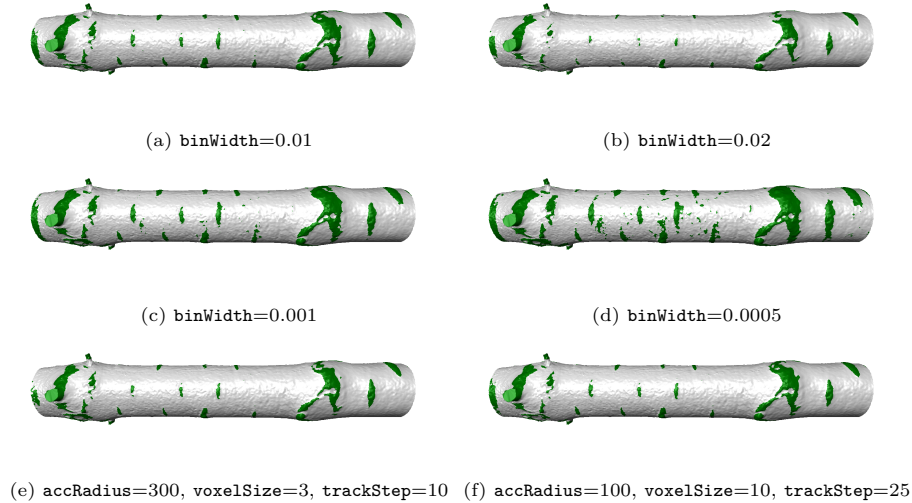
**Fig. 9.** Segmentation results of the proposed method for some logs: the detected zones, overlaying the ground truth are in yellow, non overlaying zones are in green, and the non-detected zones are in red.

**Influence of the parameters.** We experimented the influence of some parameters in order to measure their impact towards the result quality. First, we measure the influence of the patch size when the recommended criteria are not observed. In the experiments of Fig. 10, the patch size has been changed with various values and the results appear visually robust. The most significant quality variations are visible only when using not recommended values (cases of too small (resp. big) size (images (f) resp. (d) of Fig. 10) or in case of bad orientations of the patch (images (e) of Fig. 10)). In a second time, the robustness of the



**Fig. 10.** Experimentation of the method stability towards the patch size parameters (`patchWidth`, `patchHeight`).

Rosin automatic thresholding method [15] was experimented by changing its inside parameters (`binWidth`). As shown on the experiments of the figure Fig. 11, the change of this parameter is not very sensitive (see images (a-d)). Finally, the change of multiple default parameters also show small quality variations (see Fig. 11 (e,f)).



**Fig. 11.** Illustration of the influence of the `binWidth` parameter (a-d). Images (e,f) illustrate the stability by changing several parameter values.

## 4 Discussions

The previous results show that our method can precisely segment the defects and seems to be robust to different tree species or to geometrical variations. It must be preferred to the cylindrical-based method. The actual limitations appear when protruding branches are present on the log surface. To overcome this configuration, we envisage to use an additional method to segment branches before applying the proposed algorithms.

We have also observed some limit cases for the tracking algorithm used to compute the centerline which could be sensible to the voxel size and to the noisiness of the input mesh. In particular, if the voxel size is too large or if the input mesh is too noisy, during the tracking process, a point of the centerline might be located in an empty voxel (voxel with null accumulation) and the tracking process may be terminated prematurely. Even if this case can be resolved by a local search of non null accumulation, in future work we plan to use the tracking

method based on the confidence of accumulation [8]. As presented in the previous section, the choice of `voxelSize` is crucial. If the voxel size is smaller than the resolution of the data, the program may require more memory and time with no better result.

The width of bin used in the thresholding method (`binWidth`) should be small for a more precise threshold especially when one want to detect the small defects. For this reason, in our experimentations, the `binWidth` was fixed to 0.01 mm.

The choice of the patch size must guarantee that the height is several times greater than the width and enough large to avoid empty patch. Moreover the patch height should be greater than the largest defect height. In our experimentations, we fixed the width equals to 25 mm and the height equals to 100 mm.

## 5 Conclusion

This paper has presented the implementation details of our novel algorithms to segment the defects on the surface of tree logs from Terrestrial Laser Scanning data. The proposed method consists in algorithms to precisely compute the centerline of tree logs and algorithms to compute the reference distance to the centerline and to threshold the point cloud. The experiments showed that the method could precisely segment defects with complex shape like the Chinese mustaches and small defects. The actual limit case are the logs with very prominent defects like a living branch for which the proposed method did not have a good performance. All the results presented in this paper are reproducible both from the source code or from the online demonstration.

**Acknowledgment:** This work was supported by the French National Research Agency through the Laboratory of Excellence ARBRE (ANR-12- LABXARBRE-01) and by the Lorraine French Region.

## References

1. DGtal: Digital Geometry tools and algorithms library, <http://dgtal.org>
2. Online demonstration, [http://ipol-geometry.loria.fr/~kerautre/ipol\\_demo/TrunkDefaultMeasure](http://ipol-geometry.loria.fr/~kerautre/ipol_demo/TrunkDefaultMeasure)
3. Flann library (october 2016)
4. Bauer, U., Polthier, K.: Generating parametric models of tubes from laser scans. *Computer-Aided Design* 41(10), 719–729 (Oct 2009)
5. Fischler, M.A., Bolles, R.C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM* 24(6), 381–395 (1981)
6. Gough, B.: GNU scientific library reference manual. Network Theory Ltd. (2009)
7. Kerautret, B., Krähenbühl, A., Debled-Rennesson, I., Lachaud, J.O.: 3D geometric analysis of tubular objects based on surface normal accumulation. In: *Image Analysis and Processing—ICIAP 2015*, pp. 319–331. Springer (2015)

8. Kerautret, B., Krahenbül, A., Debled Rennesson, I., Lachaud, J.O.: Centerline Detection on Partial Mesh Scans by Confidence Vote in Accumulation Map (2016), to appear in the proceedings of ICPR 2016
9. Kirbas, C., Quek, F.: A review of vessel extraction techniques and algorithms. *CSUR* 36(2), 81–121 (2004)
10. Kretschmer, U., Kirchner, N., Morhart, C., Spiecker, H.: A new approach to assessing tree stem quality characteristics using terrestrial laser scans. *Silva Fenn* 47, 14 (2013)
11. Krähenbühl, A., Kerautret, B., Debled-Rennesson, I., Mothe, F., Longuetaud, F.: Knot segmentation in 3D CT images of wet wood. *Pattern Recognition* 47(12), 3852–3869 (Dec 2014)
12. Lesage, D., Angelini, E.D., Bloch, I., Funka-Lea, G.: A review of 3D vessel lumen segmentation techniques: Models, features and extraction schemes. *Medical Image Analysis* 13(6), 819–845 (Dec 2009)
13. Muja, M., Lowe, D.G.: Scalable Nearest Neighbor Algorithms for High Dimensional Data. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 36 (2014)
14. Nguyen, V.T., Kerautret, B., Debled-Rennesson, I., Colin, F., Pipoule, A., Constant, T.: Segmentation of defects on log surface from terrestrial lidar data. In: to appear in the proceedings of ICPR 2016 (2016)
15. Rosin, P.L.: Unimodal thresholding. *Pattern recognition* 34(11), 2083–2096 (2001)
16. Rusu, R.B., Cousins, S.: 3D is here: Point Cloud Library (PCL). In: *IEEE International Conference on Robotics and Automation (ICRA)*. Shanghai, China (May 9-13 2011)
17. Schütt, C., Aschoff, T., Winterhalder, D., Thies, M., Kretschmer, U., Spiecker, H.: Approaches for recognition of wood quality of standing trees based on terrestrial laserscanner data. *ISPRS* 36, 179–182 (2004)
18. Stängle, S.M., Brüchert, F., Kretschmer, U., Spiecker, H., Sauter, U.H.: Clear wood content in standing trees predicted from branch scar measurements with terrestrial lidar and verified with X-ray computed tomography 1. *Canadian Journal of Forest Research* 44(2), 145–153 (2013)
19. Thomas, L., Mili, L.: A robust gm-estimator for the automated detection of external defects on barked hardwood logs and stems. *Signal Processing, IEEE Transactions on* 55(7), 3568–3576 (2007)
20. Thomas, L., Shaffer, C.A., Mili, L., Thomas, E.: Automated detection of severe surface defects on barked hardwood logs. *Forest products journal* 57(4), 50 (2007)
21. Thomas, L., Thomas, R.E.: A graphical automated detection system to locate hardwood log surface defects using high-resolution three-dimensional laser scan data. In: *17th Central Hardwood Forest Conference*. vol. 78, p. 92 (2010)

**2.7 Article 3 : Branches and Defects Segmentation on the surface of tree trunk from partial or full terrestrial Lidar data. (Projet d'article, en cours)**



# Branches and Defects Segmentation on the surface of tree trunk from partial or full terrestrial Lidar data

Van-Tho Nguyen<sup>b</sup>, Bertrand Kerautret<sup>a</sup>, Isabelle Debled-Rennesson<sup>a</sup>, Francis Colin<sup>b</sup>, Alexandre Piboule<sup>c</sup>, Thiéry Constant<sup>b</sup>

<sup>a</sup> *Université de Lorraine, LORIA, Adagio team, UMR 7503, 54506, Nancy, France*

<sup>b</sup> *Université de Lorraine, AgroParisTech, Inra, Silva, F-54000 Nancy*

<sup>c</sup> *Pôle Recherche, Développement et Innovation de Nancy, ONF, 54 600 Villers lès Nancy, France*

---

## Abstract

Segmentation of defects on the tree log surface remains a challenge due to the slight differences of relief they induced on a bark which can be ridged furrowed, fissured, flaked, exfoliated or smooth. In previous work, we proposed a novel approach for segmenting small defects on the log surface based on the tubular object analysis. The approach showed a good result in logs of 7 species. However, its performance on logs having branches was reduced. Moreover, the proposed method to compute the centerline of tubular objects was not optimized for a large volume of data such as a mesh of an entire trunk or data describing partially a trunk and resulting from one or two scans. In this paper, we propose an extension of this work. First, we compute centerlines for multiple segments of the trunk by using the accumulation of surface normals. Second, we compute the trunk centerline by smoothing the union of segment centerlines. Third, we segment the branches. Finally, we threshold the point cloud of the trunk without branches by the difference between the distances to the centerline and the reference distance estimated from a patch of neighbors. The performance of the proposed approach was evaluated on trunks of different species. The results showed that our approach segmented branches with an accuracy of 100 %. On the same dataset, our new approach outperformed the previous approach.

---

## 1. Introduction

Tubular shapes are common in many natural or artificial objects. The segmentation and the analyze of tubular objects are often motivated by applications coming from various fields like medicine [8, 11, 21], biology [9] or industry [1, 6]. In the forestry domain, the tubular shapes are present in a variety of objects, such as trunk, branch, and knots trajectory inside woods.

The information about knots inside wood (internal defects) is an important feature for grading roundwood quality. One of the most efficient methods to detect internal defects is the computerized tomography (CT) in which X-rays are used to produce cross-sectional images of the trunk. It can detect the trace of defects inside wood with a

---

*Email addresses:* [van-tho.nguyen@inra.fr](mailto:van-tho.nguyen@inra.fr) (Van-Tho Nguyen), [bertrand.kerautret@loria.fr](mailto:bertrand.kerautret@loria.fr) (Bertrand Kerautret), [isabelle.debled-rennesson@loria.fr](mailto:isabelle.debled-rennesson@loria.fr) (Isabelle Debled-Rennesson), [francis.colin@inra.fr](mailto:francis.colin@inra.fr) (Francis Colin), [alexandre.piboule@onf.fr](mailto:alexandre.piboule@onf.fr) (Alexandre Piboule), [thiery.constant@inra.fr](mailto:thiery.constant@inra.fr) (Thiéry Constant)

diameter as small as 1 mm by manually plotting markers along the tracks of knots [3]. The segmentation and analysis of the knots using 3D volumic data recovered from CT scanner can be done manually [3] or automatically [9]. Such analysis allows biologists or sawmills to recover and measure the knot orientation and volume, useful for optimizing the industrial process of transformation or to better link timber quality to tree growth conditions [13]. While the CT can achieve a very good accuracy for defect recognition (up to 95% [12]), it has several drawbacks. Since the 3D volumic data are acquired from CT scanner, this technology is hardly applicable to measure tree wood quality on stranding tree. Moreover, the investment cost for such a system is often high and not affordable for most of the sawmills [22].

The TLS (Terrestrial Laser Scanning) appears as an alternative to CT approach in the forestry domain for grading roundwood quality. Since this technology can provide high resolution three-dimensional (3D) point cloud on an individual tree basis with millimeters precision, as a complement to the volume characterization, high point cloud densities offers a possibility to grade the standing tree quality, it leads to a new possibility to grade the standing tree quality by detecting the local defects on the trunk surfaces. Over the past decade, several studies showed that the correlation between the internal and external defects is high with coefficients of determination ( $R^2$ ) greater than 0.6 [16, 17, 20, 24]. This high correlation suggests that the external defects can be used to deduct the information about the internal defects.

The research on the segmentation of defects on trunk surfaces, however, are still scarce [10, 19, 20, 22]. These researches focused on one or more defect types related to the branching of trees. The challenge of the segmentation of defects on the trunk surfaces from TLS data is that the defects have a wide variability in shapes. The variation depends on multiple factors, such as type and dimensions of defects, and the age and bark structure of trees. The property of branching defects allowing us to detect them, is their protrusion from the regular trunk surface. Most of the approaches tried to estimate the regular surface of the trunk (the reference surface) and then use the difference between the surface and the reference surface to segment the defects. The reference surfaces were computed by using a primitive fitting such as cylinder [10, 19, 20] and circle [22, 23]. Due to the complex shape of logs like curvature, taper, ovality and the unrounded cross section these approaches can only detect the large protruding defects. For example, Thomas et al. [22] detected the branch scars with a diameter greater than 7.5 cm and protruding at least 2.2 cm from the bark. In a previous work [15], we successfully developed the detection of the surface defects which can detect defects with a diameter as small as 1 cm with a suitable spatial resolution of the 3D data on tree logs with a 1.5 m long. The approach was based on an estimation of the reference distance for each point in the point cloud to the trunk centerline from a longitudinal patch of neighbors. The centerline was computed by a modified version of the algorithm proposed by Kerautret et al. [6]. However, the method, built under the assumption of weak protrusion, has the best performance from scans of actual or virtual pruned trunks. In this context, the main purpose of this work is to complete the published work [15] to segment externally visible wood defects on trunk surfaces from TLS data acquired from standing trees including branches. The method is expected to work with trunks coming from different species, different bark roughness, log taper, ovality. In addition, we evaluate the approach with fully described portions of trunks but also with partially described portions with only one side of the trunk.

## 2. Materials and methods

### 2.1. Methodology

Beginning with a mesh recovered from a point cloud of the trunk, a subsampled point cloud is produced to accelerate the computation of the normal vectors and of the computation of the centerline. The centerline of the trunk is computed by using a modified version of the method proposed by Kerautret et al. [7]. Then, the branches are segmented by dividing the trunk volume into equal size equivalent cylindrical portions around 50 mm long and based on a rough estimation of the trunk diameter. In each sector, the closest points to the centerline belong to the trunk, other points go to the set of branch points which is removed for further processing. Next, smaller defects, such as branch scar, picot, spheroblaste, burl and bud cluster are segmented within the trunk portion as in [15]. Finally, all information about branches and defects is merged.

#### 2.1.1. Subsample of point cloud

A point cloud representing a trunk can have millions of points. In our program, the query for neighbors was used several times, such as for the estimation of normal vectors and the computation of patches. We used a KD-Tree [2] to accelerate the queries. In a 3D space, a query time is bounded by  $O(n^{\frac{2}{3}} + k)$  with  $n$  is the number of points and  $k$  is the number of returned points [5]. When we compute the normal vector for all points in the point cloud, then the complexity is  $O(nn^{\frac{2}{3}} + k)$ . This suggests that the running time increase quickly by the number of points. In a high-density point cloud, the computation of normal vectors can take a long time. We observed that a subsample of the high-density point cloud can be used to estimate the normal vectors with a reduction in execution time without a degradation in quality. Thus, we subsampled the point cloud by the following method.

- Create a voxel grid of the point cloud  $\mathcal{P}$  with a voxel size  $h$ .
- For each voxel, only keep the closet point to the center of the voxel. These points forme the subsampled point cloud  $\mathcal{S}$ .

By using this subsampling technique with a voxel size of 3 mm, which is close to the mean point spacing of the point cloud, we were able to reduce a high-density point cloud with 2.7 millions of points to a point cloud with 327 thousands of points.

#### 2.1.2. Centerline Computation

In [15], we proposed a method to compute the centerline of logs with a limited length from 1 m to 1.5 m. The method was based on the surface normal accumulation [6]. Then, a tracking procedure was used to compute the centerline. Beginning at the maximal accumulation, the tracking procedure tried to find the maximal accumulation in the next cross-section with a regular predefined step. However, in our context, this approach has two main drawbacks. First, in some case, where the next cross-section is empty, or with low values of accumulation, the maximal accumulation point does not exist or is ambiguous. Second, the tracking method needed an accumulation on the entire point cloud. Because the 3D image used to store the accumulation score must cover the bounding box of the point cloud, a point cloud of a trunk with branches has a large

volume that an ordinary computer cannot handle. We overcame the first drawback by using the confidence vote proposed by [7]. We solved the second drawback by doing the accumulation on the smaller parts of the mesh. The mesh is divided into several slightly overlapping segments. Figure 1 shows the flow to compute the centerline of the trunk.

The first step was to subsample the mesh to obtain a cloud in a smaller size to accelerate the computation of normal vectors. The normal vector of a point  $P$  in the subsampled point cloud  $S$  was estimated by using the normal of a plane fitted to its nearest neighbors. The neighbors in a radius  $R$  of  $P$  were obtained by using a range query in a KD-Tree of the subsampled point cloud. The subsampled point cloud  $S$  was then divided into smaller overlapped segments of length  $l$  mm by the z coordinate. The choice of  $l$  depends on the volume and the voxel size. The accumulation of normal vectors was computed independently for each segment as follows. From each point  $P$  with its normal vector  $\vec{n}_i$  and applying a scan along the inner normal until a maximal distance  $d_{acc}$ . During this scan, we increase by one the accumulation score of all voxels of the digitized space which are crossed by the 3D line defined from the point and the direction  $\vec{n}_i$  (see Figure 3 (a)). From this process, we obtained an accumulation map indicating for each voxel, a score corresponding to the number of normal vectors passed through it:  $v_{acc}$ . Kerautret et al. [7] proposed a novel method to increase the robustness of the centerline computation by introducing a new variable, the confidence score. The main idea was to define for each voxel the maximum occurrence score  $v_{max}$ , after the calculation of the accumulation score  $v_{acc}$ . For a voxel  $v$ , the value  $v_{max}$  is defined as the number of normal vectors crossing through  $v$  and for which the accumulation note  $v_{acc}$  is maximum along the scan direction. The confidence score of a voxel ( $v_{conf}$ ) of the accumulation image has a value in the range of  $[0, 1]$  and is defined by Equation 1. Figure 3 illustrates the computation of  $v_{conf}$ .

$$v_{conf} = \frac{v_{max}}{v_{acc}} \quad (1)$$

We utilized a simple approach that threshold the confidence scores of each segment after the accumulation. The histogram of the confidence scores had a unimodal distribution, consequently, an automatic thresholding using the Rosin’s method [18] on the histogram of the confidence scores allows us to obtain the voxels in the accumulation image that have the highest scores (Figure 4). During the accumulation process, we also computed for each voxel the mean of the distances from points  $P_s$  to the voxel  $v$  denoted  $v_r$ , with  $P_s$  the set of points, whose normal vector passed through  $v$ . The thresholding result contains some point even with a high confidence score but corresponds to the other tubular shapes such as branches or imperfections of trunk, a post-processing step was used to remove these voxels. The filter was  $v_r \notin [-0.05 + v_m, 0.05 + v_m]$  where  $v_m$  was the median of  $v_r$  of the segment. The threshold result was the voxels in the vicinity of the segment center. The centerline of the trunk was obtained by an interpolation using BSpline [4] on the union set of all the segment center points.

### 2.1.3. Branch segmentation

The patch-based method for segmenting the defects that was introduced in the previous work [15] required that there was no presence of long branches on the trunk. To overcome this drawback, we proposed a method to segment branches from the trunk

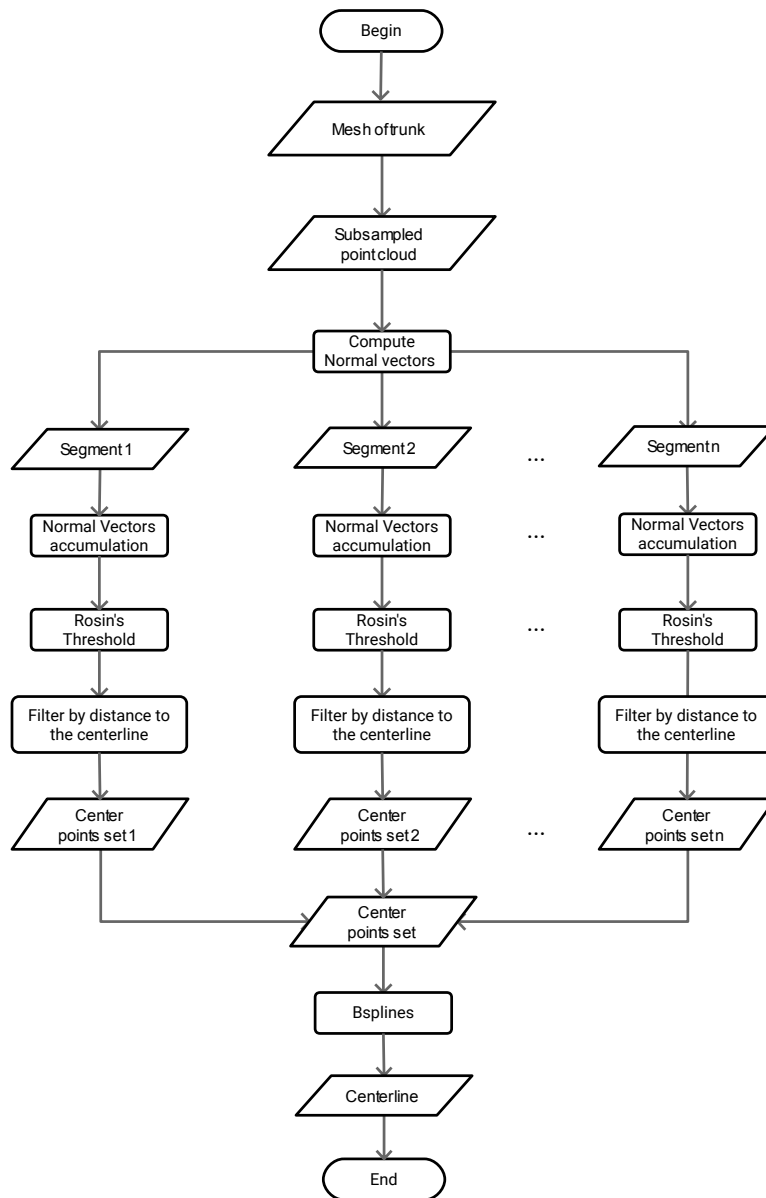


Figure 1: Algorithm for computing of the centerline.



Figure 2: Illustration of the trunk centerline computation using an accumulation of normal vectors using 4 overlapped segments. In darkest grey were the overlapped regions.

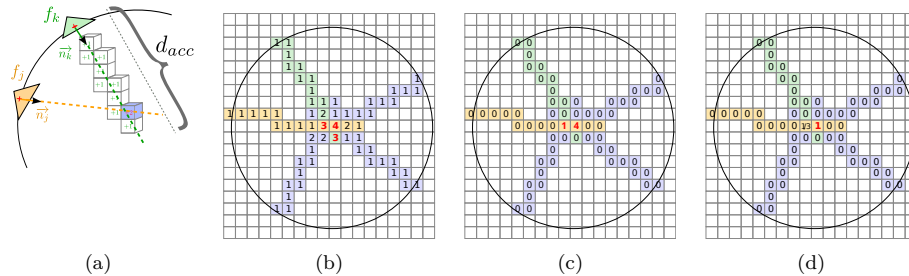


Figure 3: Computation of the confidence score [7]. (a) Scan from mesh elements along a distance  $d_{acc}$ , (b) accumulation score  $v_{acc}$ , (c) maximum occurrence score  $v_{max}$  and (d) confidence score  $v_{conf}$ .

before using the patch-based method. The idea of the branches segmentation is to separate the points into two disjointed sets: (1) the set  $\mathcal{T}$  contains the closer points to the trunk surface, and (2) the set  $\mathcal{B}$  contains the branches. If we consider the trunk as a tubular object, and divided the tubular object into angular sectors based on a set of short cylinders following the centerline, and with diameters based on the data, then the closest point to the centerline will have the highest probability to belong to the trunk. At this step, we had the set  $\mathcal{T}$  that contains the closest point to the centerline of all the sectors and the set  $\mathcal{B}$  contains all others points. Because in each sector, we add only one point to the set  $\mathcal{T}$ , there was a high probability that in the set  $\mathcal{B}$ , there are points that belong to the trunk. The next step was to take back these points from the  $\mathcal{B}$  to the set  $\mathcal{T}$  using a radius search. The branches segmentation algorithm is described as follows:

- Estimation of the trunk radius  $r_m$ , using the mode of the distance to the centerline of all points in the point cloud.
- Division of the point cloud volume into slices with a thickness of  $l$  mm, following the centerline direction. Each slice is then divided into cylindrical sectors with an angle of  $\frac{l}{r_m}$  radian. In our experiment, the  $l$  value was set to 50 mm.
- For each cylindrical sector, the nearest point to the trunk centerline is added to the set  $\mathcal{T}$ , the other points of the portion are added to the set  $\mathcal{B}$ .
- For each point  $P$  in the set  $\mathcal{T}$ , we find the subset  $S$  of the set  $\mathcal{B}$ , such as the distance

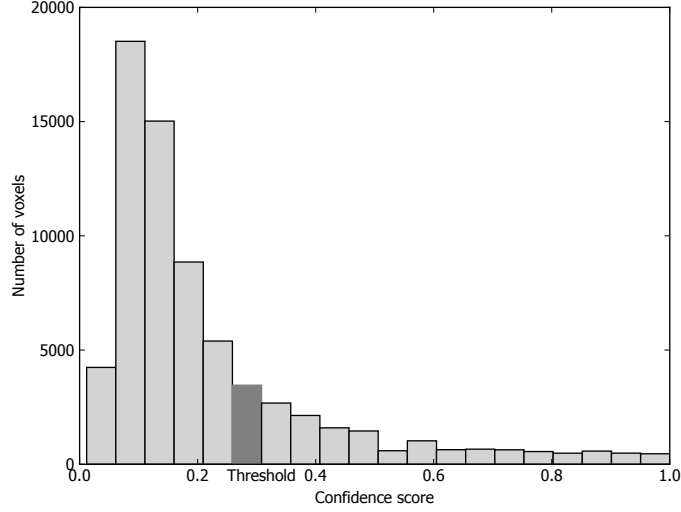


Figure 4: Automatic thresholding using Rosin’s method on the confidence scores.

between a point  $S_i \in S$  and  $P$  is less than or equal to  $\sqrt{2}l$ . Then, we add the set  $S$  to the set  $\mathcal{T}$  and remove  $S$  from the set  $\mathcal{B}$ :  $S = \{b \in \mathcal{B} | \text{distance}(b, P) \leq \sqrt{2}l\}$   
 $\mathcal{T} = \mathcal{T} \cup S$  and  $\mathcal{B} = \mathcal{B} \setminus S$ .

#### 2.1.4. Compute the reference distance to the centerline

The reference distance to the centerline of a point is defined as the distance from the point to the centerline with a hypothesis that the trunk is defect-free. In other words, if a point does not belong to a defect the distance to the centerline and the reference distance to the centerline should be close. In contrast, if a point belongs to a defect, the reference distance to the centerline should be smaller than the distance to the centerline. At this step, we had two sets, the set  $\mathcal{T}$  containing points belong to the trunk and the set  $\mathcal{B}$  containing the points belonging to branches. The computation of the reference distance to the centerline is for points in the set  $\mathcal{T}$ . To compute the reference distance to the centerline, first, we represent the set  $\mathcal{T}$  in the cylindrical coordinate system  $(r, \theta, z)$  based on the centerline. The coordinates of a point  $P$  are computed as follows (see [15] for illustrations of the conversion).

- $r$  is the distance from  $P$  to the centerline.
- $z$  is the height of the point along the centerline.
- $\theta$  is the angle between the radial vector and the  $Oy$  axis of local coordinates of the segment.

For a more robust method and to reduce the running time, we use another subsampled point set to compute the reference distance. The density of the point cloud is often higher on defects, which leads to a bias toward it when doing a regression (Figure 7 (a)). The idea is the same than we do to segment branches. That is the division of the trunk

volume into equal size cylindrical sectors. However, in this procedure, we keep only the closest point to the centerline within a smaller sector with the arc length at trunk radius relatively equal to the mean of point spacings of the point cloud (3 mm in our case). The subsampled of the set  $\mathcal{T}$  denoted  $\mathcal{C}$  contains the closest points to the centerline.

Second, for each point  $P(r, \theta, z)$  of the set  $\mathcal{T}$  in the cylindrical coordinate system, we compute a narrow patch  $\mathcal{P}_P$  containing the neighbors of  $P$  in the subset  $\mathcal{C}$  of  $P$  by the equation 2.

$$\mathcal{P}_P = \{Q \in \mathcal{C} : |Q_\theta - P_\theta| \leq \frac{\varphi}{2}, |Q_z - P_z| \leq \frac{h}{2}\} \quad (2)$$

with  $\varphi = \frac{\text{patch width}}{R_m}$ ,  $h$  is the patch height and  $R_m$  is the mode of the distances to the centerline.

For each point  $P(r, \theta, z)$  in the point cloud, the reference point  $\hat{P}(r, \theta, z)$  is computed from  $P$  and  $\mathcal{P}_P$  with  $\hat{P}_\theta = P_\theta$  and  $\hat{P}_z = P_z$ . The coordinate  $\hat{P}_r$  (Equation 4) is estimated from a linear regression (Equation 3) established from  $P_z$  and the distance to the centerline for all points belonging to  $\mathcal{P}_P$ .

$$r = az + b \quad (3)$$

$$\hat{P}_r = aP_z + b \quad (4)$$

#### 2.1.5. Thresholding

First, for each point  $P \in \mathcal{T}$  we compute the difference between the distance to the centerline  $P_r$  and its reference distance to the centerline  $\hat{P}_r$ .  $\delta P_r = P_r - \hat{P}_r$ . As presented in [15], the histogram of the differences is unimodal, and the Rosin's method [18] give the optimized threshold. This method designed to work with a unimodal histogram, in which a straight line is drawn from the maximum to the first empty bin of the histogram, to form a triangular-like shape. The threshold  $T$  is selected at the point of the histogram that maximizes the perpendicular distance from the histogram to the straight line. The point cloud is then thresholded by this threshold  $T$ , with defect points having a difference between the distance to the centerline and the reference distance to the centerline greater than  $T$  (Equation 5).

$$g(P) = \begin{cases} 1 & \text{if } \delta(P_r) > T \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

## 2.2. Material and evaluation criteria

In order to evaluate the proposed method, we carried out three experiments on the mesh of different trunks and logs. In the first experiments, we evaluate the centerline computation and the branch segmentation of a trunk recovered from a European beech and a half trunk recovered from an oak. Each mesh is the data of a trunk from 0.5 m above ground to 5 or 7 m. These trees have many branches. In these tests, we did not evaluate the performance on smaller defects segmentations. The second test was on an oak trunk from 0.5 m above ground to 6 m. In this test, we compared the performance of the segmentation on the mesh built from one scan with the performance on the mesh built with 4 scans using a visualization of detected defects on meshes, the precision (Equation 6), and recall (Equation 7) criteria. The third experiment, we retested the 10



logs from 7 different species, which were tested in previous work [15]. In this database, we created a ground truth by a manual segmentation of the defects on the meshes. The results of the automatic segmentation were compared to the ground truth at a point level and statistical measured by using the criteria such as precision, recall, and  $F_1$  (Equation 8) score.

$$P = \frac{TP}{TP + FP} \quad (6)$$

$$R = \frac{TP}{TP + FN} \quad (7)$$

where  $TP, FP, FN$  are true positive, false positive, and false negative respectively. Their definition is as follow.

- $TP$  is the number of points correctly classified as in a defect.
- $FP$  is the number of points incorrectly identified as in a defect.
- $FN$  is the number of points incorrectly identified as in a non-defect.

$$F_1 = 2 \frac{P \cdot R}{P + R} \quad (8)$$

### 3. Results and discussions

#### 3.1. Centerline computation

The centerline computation results were robust to the trunk species, partial mesh and the presence of branches. The computation of the centerline was robust to the presence of branches (Figure 5 (a)) and to the data of the half side of the trunk (Figure 6 (a)). Figure 5 shows a European beech with a strongly curved trunk, which had many branches. This trunk also had an imperfect round cross-section. Figure 6 presents the mesh of an oak, made from only one TLS scan.

#### 3.2. Branch segmentation and defect segmentation

As showed in Figure 5 (b) and Figure 6 (b), all the branches were segmented. As expected, there were small portions of each branch remaining attached to the trunk after the branch segmentation. These portions were segmented by our defect segmentation algorithm (Figure 5 and Figure 6 (c)). Figure 5 (c) and Figure 6 (c) show that the algorithm can detect both branches and defects. Figure 7 shows a comparison of the profiles of the distances to the centerline (a) between the original point cloud and the subsample point cloud (b). The result on the subsample patch was better than the original patch.

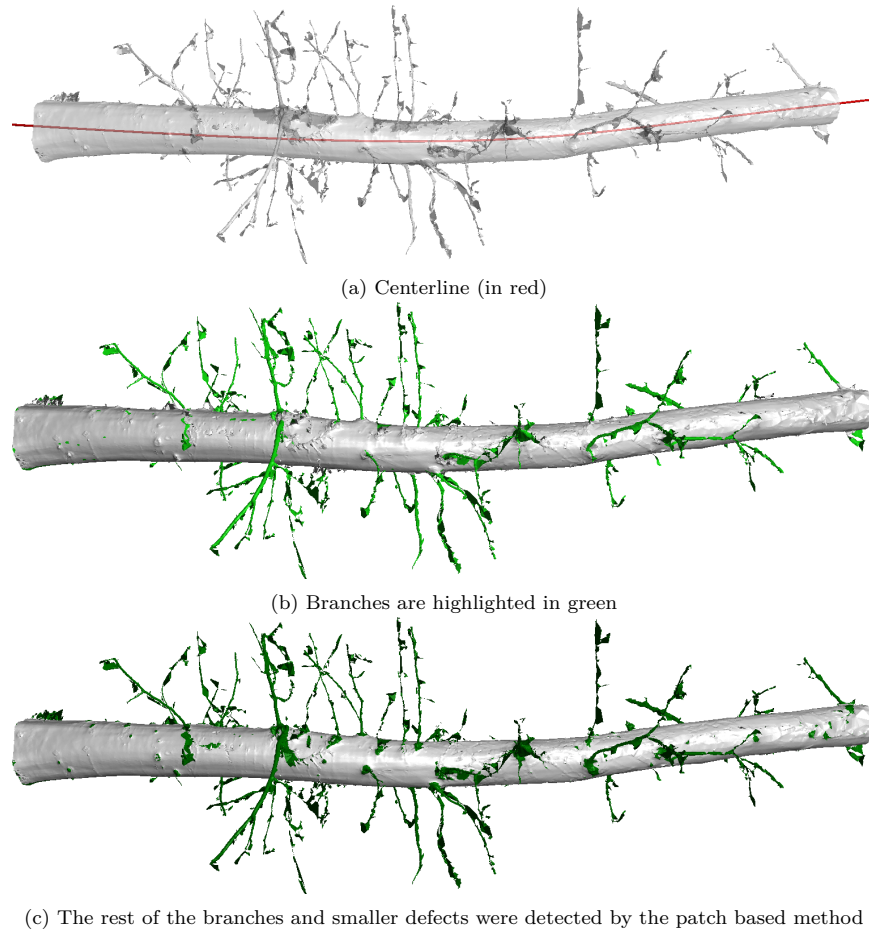


Figure 5: Segmentation results of the proposed method on an European beech, which has a curve trunk and many branches.

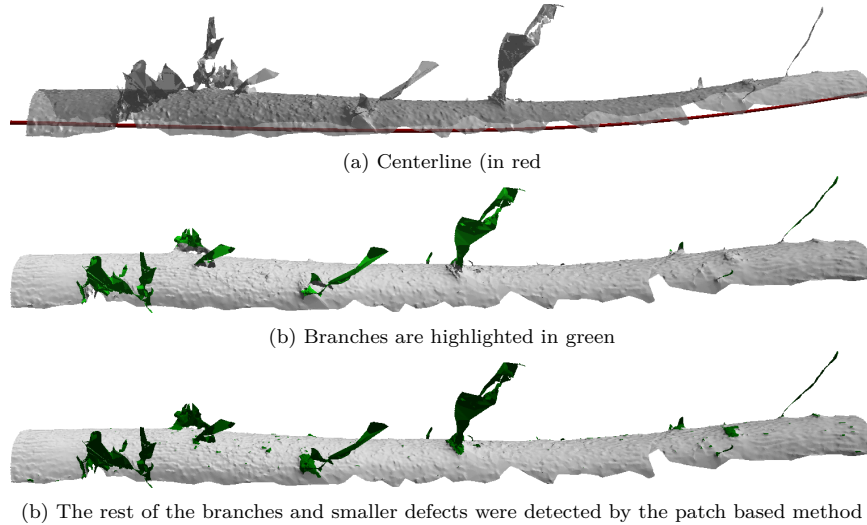


Figure 6: Segmentation results of the proposed method on a partial trunk of an oak.

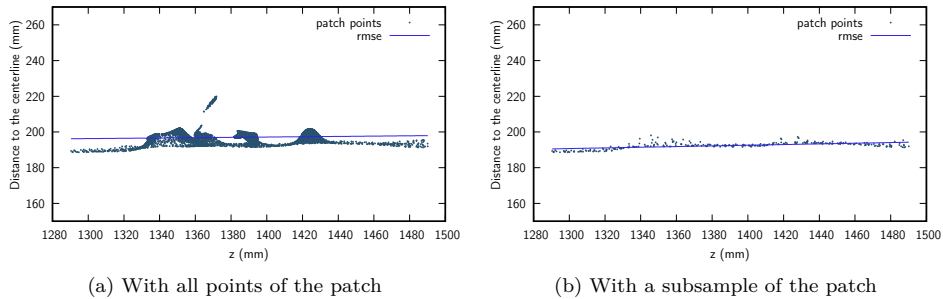


Figure 7: Line fitting to the profile of the distances to the centerline of a patch.

### 3.3. Segmentation of defects on the mesh of partial trunk

Thanks to the ability to compute the centerline even in the mesh built from only one scan of the trunk. As shown in Figure 8, the performance of the segmentation algorithm on these meshes was not far from the mesh of the entire trunk, which was scanned from 4 different points of view. The weak point when using just one scan is that the spatial resolution of the border of such a scan is lower than in the center area and there are more borders leading to fewer points in the corresponding cylindrical sectors. Thus, the information about the defects was less precise compared to 4 scans (see Figure 8 (a) and (b) for example). The Table 1 also reflect this finding as the precision is very high (0.910) while the recall is lower (0.659). The multiscan can also benefit from the reduction of occlusions (Figure 8 (g) and (h)). While in the one scan mesh, there are several large missing zones, on the 4 scans mesh, these zones are compensated by points from other

scans.

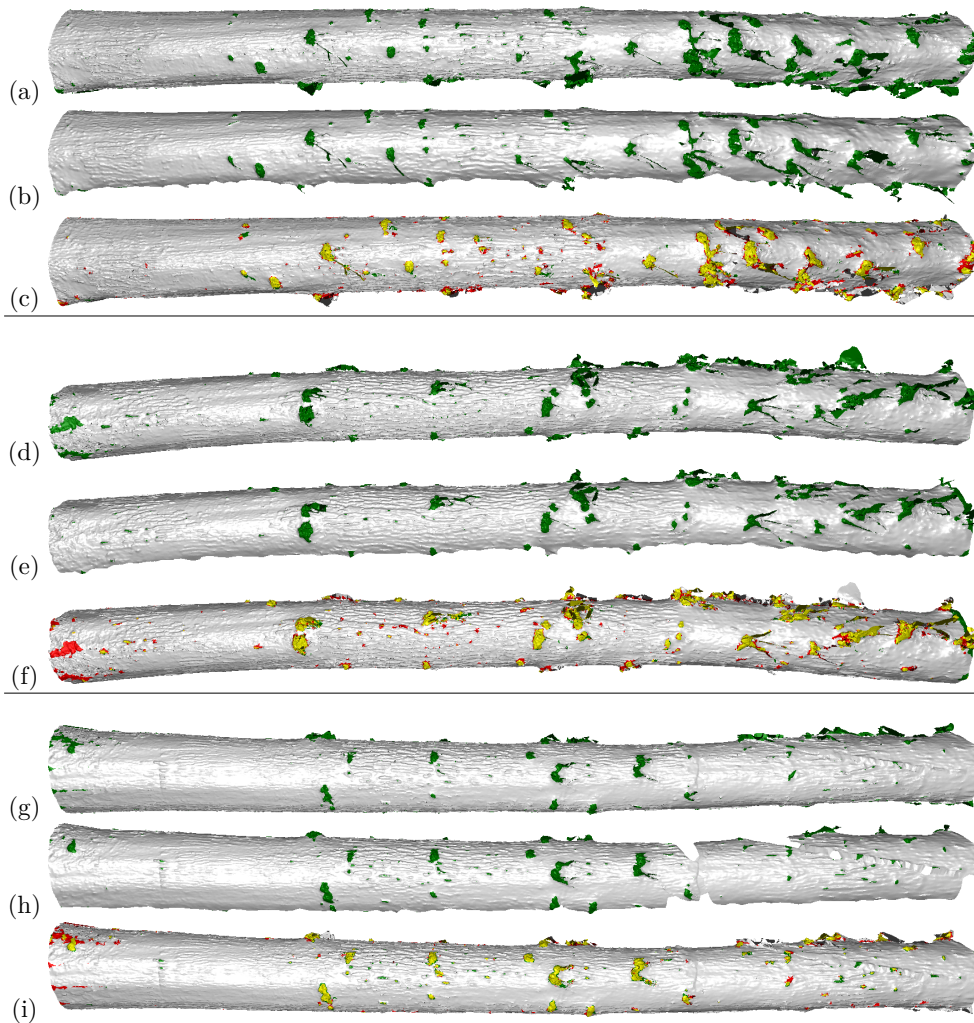


Figure 8: Comparison of results between a mesh built from 4 scans (a, d, g) and meshes built from only one scan (partial trunk) of the same trunk (b, e, h). The figures (c, f, i) show the superimposition of the detected defects areas on the mesh of partial trunk and on the entire trunk. The zones detected in both meshes are in yellow, the zones detected only on the mesh of entire trunk are in red, and the zones detected only on the mesh of the partial trunk (one scan) are in green.

#### 3.4. Comparison with previous work

Table 2 shows the segmentation results on the database consisting of 10 logs that were tested in Nguyen et al. [15]. In comparison with the results of [15], we see an improvement

Table 1: Comparison of the results on meshes built from one scan and the mesh built from 4 scans of the same oak trunk using the same parameters values: patch width = 40, patch height = 200 and bin width = 0.01. The result on the 4 scans mesh was used as the reference.

Scan	Precision	Recall	F1
face 1	0.923	0.634	0.751
face 2	0.953	0.683	0.794
face 3	0.889	0.627	0.734
face 4	0.821	0.689	0.748
Average	0.910	0.659	0.763

in the overall  $F_1$  score from 0.710 to 0.76. Most logs had an improvement. We see the improvement when there was the presence of branch tips on the logs. In particular, on the log of species that have smooth bark, such as beech, wild cherry tree, and red oak, there is often a deformation at the intersection between the branch and the trunk. The shape of the deformation is often denoted as "Chinese mustache" or "eyebrows". In our previous work [15], the algorithm could not segment all these parts and often resulted in an over-segmentation. The new approach, by segmenting the branch before the segmentation of smaller defects erases this kind of difficulties to segment these type of defects (Figure 9).

Table 2: Result of the segmentation algorithm of the proposed method and [15] using the common parameters values: patch width = 25, patch height = 100 and bin width = 0.01

Log	Proposed method			method[15]		
	Prec <sup>1</sup>	Recall	F1	Prec	Recall	F1
Fir 1	0.672	<b>0.796</b>	0.727	<b>0.747</b>	0.769	<b>0.757</b>
Fir 2	0.659	<b>0.815</b>	<b>0.728</b>	<b>0.673</b>	0.775	0.719
Wild cherry 1	<b>0.745</b>	<b>0.853</b>	<b>0.794</b>	0.696	0.765	0.728
Wild cherry 2	<b>0.865</b>	<b>0.894</b>	<b>0.878</b>	0.846	0.711	0.771
Red oak 1	0.724	<b>0.863</b>	<b>0.786</b>	<b>0.749</b>	0.742	0.744
Red oak 2	<b>0.489</b>	0.760	<b>0.594</b>	0.428	<b>0.833</b>	0.564
Beech	0.662	<b>0.694</b>	<b>0.676</b>	<b>0.670</b>	0.604	0.634
Birch	0.694	<b>0.784</b>	0.736	<b>0.733</b>	0.756	<b>0.744</b>
Elm	<b>0.706</b>	<b>0.764</b>	<b>0.733</b>	0.694	0.755	0.721
Wild service tree	<b>0.338</b>	<b>0.798</b>	<b>0.473</b>	0.247	0.741	0.370
Overall	<b>0.705</b>	<b>0.824</b>	<b>0.758</b>	0.685	0.740	0.710

<sup>1</sup> Precision

### 3.5. Influence of parameters

The experiments used several parameters. The parameters that have the most impact on the performance of the algorithms are the voxel size, the bin width of the histogram using to find threshold by using the Rosin's method [18] and the patch size. In Nguyen et al. [14], we discussed in detail of the influence of the patch size and the bin width. In addition to these parameters, our new algorithm requires other parameters such as the size of cylinder sectors used to segment branches. The set of parameters must result in the best branch segmentation between its distal and proximal components, with the later containing bark irregularities and noise. The cylinder length must be long enough

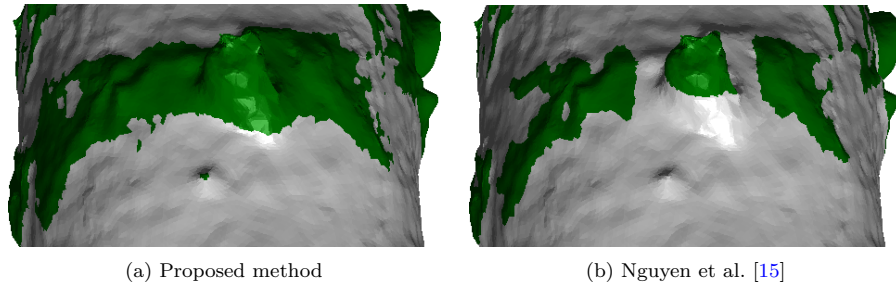


Figure 9: The improvement in segmentation performance when there is a presence of a branch tip on a wild cherry tree.

to avoid the complete inclusion of a cylindrical sector in a big branch, with a setting to 50mm this condition is a priori respected. A small value of  $l$  is not recommended since it can lead to portions of branches identified as a part of trunk.

The voxel size is also an important parameter. It has an impact on the execution time and the precision of the centerline computation. In general, the smaller voxel size the more precise the centerline but requires a more memory and a longer execution time. The optimized voxel size is related to the resolution of the point cloud, as if it is smaller than the point spacing, the result would not be improved with an expense of memory usage and execution time.

The number of segments indicates how many segments that the trunk will be divided into for computing the centerline. The choice of this parameter is mainly depending on the available memory with respect to the choice of the voxel size, the volume of the trunk. A small number of this parameter can lead to an excessively requirement of memory that the running computer can not offer. A large number can lead to a less accurate of the computed centerline because the segment length can be too small. We suggest that the segment length should be several times greater than the point spacing.

As discussed in [14], the patch size is one of the most important parameters that influence the results of the segmentation. However, the patch height should be greater than then patch width [14]. From our experimentation, the choice of the patch size was depending on firstly the roughness of the bark and the defect sizes. On the smooth bark species, a smaller patch size often allows detecting small defects. A good value for patch width was between 20 mm and 40 mm and patch height was between 100 mm and 200 mm. On furrowed bark trunks, a good patch width was between 40 mm and 80 mm and a good patch height was between 200 mm and 400 mm. The patch height should also be chosen with respect to the defect dimensions to avoid a case with a large portion of a patch belonging to a defect.

#### 4. Conclusion

This paper proposed an improvement of a previous approach for segmenting the defects on the surface of tree logs described with Terrestrial laser Scanning data a novel approach for segmenting the defects on the surface of tree logs from Terrestrial Laser Scanning data. This work completed our previously published method by including the

branch segmentation in the processing workflow, adaptations to process big volumes of 3D data or a partial trunk description with only one scan, and an improvement of the reference distance calculation proposed in [15].

The experiments showed that the method could precisely segment defects with a complex shape like the Chinese mustaches and small defects whatever the shape of the trunk is and with very few parameters easy to define with reference to the type of bark, and the class size of branches diameters. Moreover, by comparing with a method based on cylindrical fitting, the results were more efficient for different tree species and for logs deviating from the cylindrical model.

### Acknowledgments

This work was supported by the French National Research Agency through the Laboratory of Excellence ARBRE (ANR-12- LABXARBRE-01) and by the Lorraine French Region.

### References

- [1] Bauer, U., Polthier, K., 2009. Generating parametric models of tubes from laser scans. *Computer-Aided Design* 41, 719–729. doi:[10.1016/j.cad.2009.01.002](https://doi.org/10.1016/j.cad.2009.01.002).
- [2] Bentley, J.L., 1975. Multidimensional binary search trees used for associative searching. *Communications of the ACM* 18, 509–517.
- [3] Colin, F., Mothe, F., Freyburger, C., Morisset, J.B., Leban, J.M., Fontaine, F., 2010. Tracking rameal traces in sessile oak trunks with X-ray computer tomography: Biological bases, preliminary results and perspectives. *Trees - Structure and Function* 24, 953–967. doi:[10.1007/s00468-010-0466-1](https://doi.org/10.1007/s00468-010-0466-1).
- [4] Gough, B., 2009. GNU scientific library reference manual. Network Theory Ltd.
- [5] Kakde, H.M., 2005. Range searching using kd tree. Florida State University .
- [6] Kerautret, B., Krähenbühl, A., Debled-Rennesson, I., Lachaud, J.O., 2015. 3d geometric analysis of tubular objects based on surface normal accumulation, in: *Image Analysis and Processing—ICIAP 2015*. Springer, pp. 319–331.
- [7] Kerautret, B., Krähenbühl, A., Debled-Rennesson, I., Lachaud, J.O., 2016. Centerline detection on partial mesh scans by confidence vote in accumulation map, in: *Pattern Recognition (ICPR), 2016 23rd International Conference on, IEEE*. pp. 1376–1381.
- [8] Kirbas, C., Quek, F., 2004. A review of vessel extraction techniques and algorithms. *CSUR* 36, 81–121.
- [9] Krähenbühl, A., Kerautret, B., Debled-Rennesson, I., Mothe, F., Longuetaud, F., 2014. Knot Segmentation in 3d CT Images of Wet Wood. *Pattern Recognition* .
- [10] Kretschmer, U., Kirchner, N., Morhart, C., Spiecker, H., 2013. A new approach to assessing tree stem quality characteristics using terrestrial laser scans. *Silva Fenn* 47, 14.
- [11] Lesage, D., Angelini, E.D., Bloch, I., Funka-Lea, G., 2009. A review of 3d vessel lumen segmentation techniques: Models, features and extraction schemes. *Medical Image Analysis* 13, 819–845. doi:[10.1016/j.media.2009.07.011](https://doi.org/10.1016/j.media.2009.07.011).
- [12] Li, P.L.P., a.L. Abbott, Schmoltdt, D., 1996. Automated analysis of CT images for the inspection of hardwood logs, in: *Proceedings of International Conference on Neural Networks (ICNN'96), IEEE*. pp. 1744–1749. doi:[10.1109/ICNN.1996.549164](https://doi.org/10.1109/ICNN.1996.549164).
- [13] Longuetaud, F., Mothe, F., Kerautret, B., Krähenbü, A., Hory, L., Leban, J.M., Debled-Rennesson, I., 2012. Automatic knot detection and measurements from X-ray CT images of wood: A review and validation of an improved algorithm on softwood samples. *COMPAG* 85, 77–89.
- [14] Nguyen, V.T., Kerautret, B., Debled-Rennesson, I., Colin, F., Piboule, A., Constant, T., 2016a. Algorithms and implementation for segmenting tree log surface defects, in: *International Workshop on Reproducible Research in Pattern Recognition, Springer*. pp. 150–166.

- [15] Nguyen, V.T., Kerautret, B., Debled-Rennesson, I., Colin, F., Piboule, A., Constant, T., 2016b. Segmentation of defects on log surface from terrestrial lidar data, in: Pattern Recognition (ICPR), 2016 23rd International Conference on, IEEE. pp. 3168–3173.
- [16] Pyörälä, J., Kankare, V., Vastaranta, M., Rikala, J., Holopainen, M., Sipi, M., Hyypä, J., Uusitalo, J., 2017. Comparison of terrestrial laser scanning and x-ray scanning in measuring scots pine (*pinus sylvestris* l.) branch structure. *Scandinavian Journal of Forest Research* , 1–8.
- [17] Racko, V., 2013. Verify the accuracy of estimation the model between dimensional characteristics of branch scar and the location of the knot in the beech trunk. *Annals of Warsaw Univ. of Life Sciences-SGGW. Forestry & Wood Technology* 84.
- [18] Rosin, P.L., 2001. Unimodal thresholding. *Pattern recognition* 34, 2083–2096.
- [19] Schütt, C., Aschoff, T., Winterhalder, D., Thies, M., Kretschmer, U., Spiecker, H., 2004. Approaches for recognition of wood quality of standing trees based on terrestrial laserscanner data. Laser-scanners for forest and landscape assessment. Proceedings of the ISPRS working group VIII/2, Freiburg, Germany. *International Archives of Photogrammetry, Remote Sensing, and Spatial Information Sciences* 36, 179–182.
- [20] Stängle, S.M., Brüchert, F., Kretschmer, U., Spiecker, H., Sauter, U.H., 2013. Clear wood content in standing trees predicted from branch scar measurements with terrestrial lidar and verified with x-ray computed tomography 1. *Canadian Journal of Forest Research* 44, 145–153.
- [21] Tankyevych, O., Talbot, H., Passat, N., Musacchio, M., Lagneau, M., 2011. Angiographic image analysis, in: *Medical Image Processing*. Springer, pp. 115–144.
- [22] Thomas, L., Shaffer, C.A., Mili, L., Thomas, E., 2007. Automated detection of severe surface defects on barked hardwood logs. *Forest products journal* 57, 50.
- [23] Thomas, L., Thomas, E., et al., 2010. A graphical automated detection system to locate hardwood log surface defects using high-resolution three-dimensional laser scan data, in: *Proceedings of the 17th Central Hardwood Forest Conference*, pp. 5–7.
- [24] Thomas, R.E., 2009. Modeling the relationships among internal defect features and external apalachian hardwood log defect indicators. *Silva Fennica* 43, 447–456.





## Chapitre 3

# Identification des défauts

### 3.1 Contexte et objectifs

L'évaluation de la qualité des bois ronds ou des arbres sur pied comprend non seulement la détection de défauts, mais aussi l'identification de leur type. Une fois les défauts détectés par les algorithmes de segmentation présentés dans le chapitre précédent, il faut identifier le type de ces défauts et veiller à supprimer les fausses détections. L'identification d'un défaut est liée à la reconnaissance de forme. Ceci constitue un défi majeur dans le domaine de la vision par ordinateur. En effet, cela est rendu particulièrement difficile à cause de la grande variété de forme de ces objets biologiques. Dans notre cas, cette variabilité tient à plusieurs facteurs notamment l'espèce de l'arbre, son type d'écorce, l'âge du défaut et toute l'histoire de son développement en lien avec l'environnement de l'arbre.

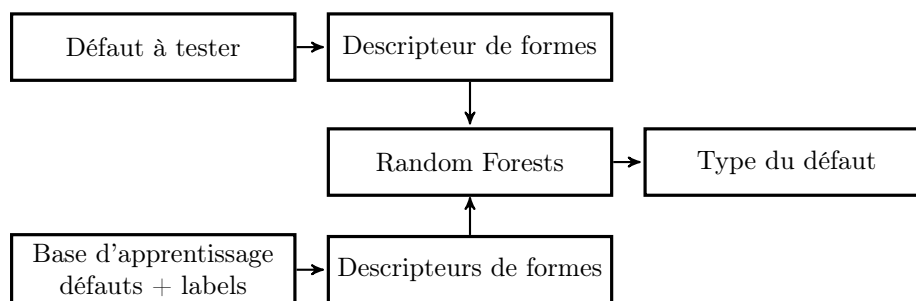
Il existe une grande diversité de méthodes dans le domaine de la reconnaissance de forme qui peuvent être utilisées dans cette partie. La plupart des approches sont basées soit sur des modèles paramétriques (Terzopoulos et Fleischer, 1988; Devroye et al., 2013), soit sur des techniques d'apprentissage automatique (Cortes et Vapnik, 1995; Bishop, 1995; Breiman, 2001). Ces dernières présentent des avantages pour décrire des objets complexes avec une grande variabilité de forme. Il a été prouvé qu'elles étaient plus efficaces que les modèles paramétriques selon Niemeyer et al. (2014). En particulier, Random Forests (Breiman, 2001) est une méthode d'apprentissage automatique supervisée basée sur des ensembles d'arbres de classification (Breiman et al., 1984). Random Forests possède de nombreuses propriétés intéressantes, telles qu'une haute précision, une bonne robustesse contre le sur-apprentissage, le bruit ou le manque de données dans l'ensemble d'apprentissage. Random Forests est utilisé dans de nombreuses recherches dans le domaine forestier et de la télédétection. Il permet de classer la couverture du sol (Chehata et al., 2009; Guo et al., 2011) en utilisant le LiDAR aérien, d'évaluer la biomasse (Mutanga et al., 2012) et également d'identifier l'espèce des arbres (Othmani et al., 2013) à partir de données TLS.

Dans le chapitre précédent, nous avons présenté une approche pour détecter les défauts à la surface du tronc à partir de données TLS. Les résultats des expérimentations étaient prometteurs avec un fort accord entre le résultat de la détection et les données de référence. Les résultats contenaient cependant de nombreux faux positifs, comme conséquence de l'objectif de ne pas oublier de défauts. La contre-partie est la nécessité de réaliser une analyse plus détaillée destinée à typer les défauts et réduire le nombre des faux positifs.

Dans ce contexte, l'objectif de ce chapitre est de classer les défauts détectés sur la surface du tronc par des algorithmes développés dans le chapitre précédent. Les types de défauts sont la branche, la cicatrice de branche, le broussin, le sphéroblaste, l'amas de bourgeons et le picot. Notre objectif particulier est de développer une méthode générique qui prend en charge les types et les formes de défaut sur les essences les plus commercialisées, en particulier le chêne et le hêtre qui ont des caractéristiques très différentes notamment la rugosité de leur écorce. D'autres espèces ont également été testées telles que le sapin, le pin, l'épicéa qui ont des structures d'écorce encore différentes. Ce chapitre fait l'objet d'une publication (section 3.5).

## 3.2 Méthodologie

Dans cette étude, nous avons choisi le classifieur Random Forests pour les raisons évoquées précédemment. Il s'agit d'une méthode d'apprentissage automatique supervisée (Figure 3.1) de type "ensemble learning" qui agrège des arbres de classification et de régression (CART) (Breiman et al., 1984).



**Figure 3.1** Schéma du principe d'une méthode d'apprentissage automatique supervisée

Dans cette étude, Random Forests a été entraîné avec 425 défauts qui ont été manuellement ou automatiquement segmentés par l'algorithme présenté dans le chapitre précédent. Le vecteur caractéristique utilisé par Random Forests comprenait les caractéristiques suivantes :

- 1) L'espèce de l'arbre.
- 2) Le ratio entre le nombre de points du défaut et le volume de sa boîte englobante.
- 3) La largeur du défaut.
- 4) Le ratio entre la largeur et la hauteur du défaut.

- 5) Le ratio entre la largeur et l'épaisseur du défaut.
- 6) La moyenne de la différence entre la distance à la ligne centrale et la distance de référence de tous les points du défaut.
- 7) L'écart-type de la différence entre la distance à la ligne centrale et la distance de référence de tous les points du défaut.
- 8) Les moments invariants de Hu.
- 9) Le ratio entre les valeurs propres  $\lambda_1$  et  $\lambda_3$ .
- 10) Le ratio entre les valeurs propres  $\lambda_2$  et  $\lambda_3$ .
- 11) L'angle entre le vecteur propre  $\vec{v}_3$  et la ligne centrale du tronc.

Où  $\lambda_1$ ,  $\lambda_2$  et  $\lambda_3$  sont les valeurs propres correspondant aux vecteurs propres du défaut  $\vec{v}_1$ ,  $\vec{v}_2$  et  $\vec{v}_3$  respectivement.  $\vec{v}_1$  est le vecteur propre ayant le plus petit angle avec le vecteur radial du tronc au centre de l'intersection entre le défaut et le tronc.  $\vec{v}_2$  est le vecteur propre ayant le plus petit angle avec le vecteur tangentiel du tronc au centre de l'intersection entre le défaut et le tronc.  $\vec{v}_3$  est le vecteur propre ayant le plus petit angle avec la ligne centrale du tronc au hauteur du défaut.

La base d'apprentissage incluait les classes, i.e. types de défaut, suivants :

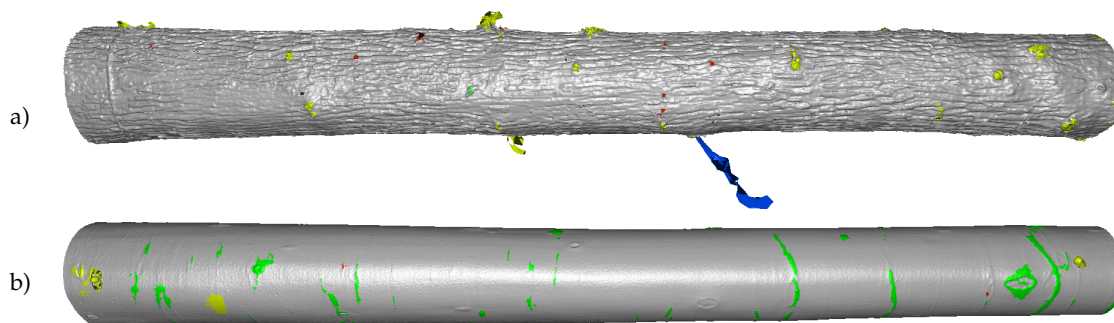
- Branche incluant branche séquentielle et branche épécormique,
- Cicatrice de branche,
- Broussin,
- Petits défauts incluant picot, sphéroblaste, amas de bourgeons,
- Écorce, classe créée pour recueillir beaucoup de faux positifs liés à une écorce un peu plus rugueuse que la normale sans que cela soit réellement un défaut.

Le principe du classifieur Random Forests repose sur un vote majoritaire. Pour chaque défaut candidat (dont la classe est à déterminer), un vecteur de caractéristiques est calculé puis transmis au classifieur Random Forests. Cela consiste en un ensemble d'arbres de classification et de régression (Breiman et al., 1984). Chaque arbre procède à une classification, ou en d'autres termes, donne un vote pour une classe. Random Forests rassemble ces votes et choisit la classe votée par la majorité des arbres.

### 3.3 Résultats principaux

Afin d'évaluer la performance de la classification, nous avons mené une expérimentation sur 26 troncs d'arbres de 6 espèces chêne, hêtre, merisier, pin, sapin et épicéa. Le Random Forests a été entraîné avec 425 défauts identifiés sur 16 troncs de 5 espèces. La base de test comprend 185 défauts présents sur 10 troncs de 6 espèces donc un tronc d'épicéa, espèce absente de la base utilisée pour entraîner le Random Forests. Dans cette expérimentation nous avons comparé le résultat de la classification avec des mesures observées de référence donnant le type du défaut. La Figure 3.2 montre des visualisations de résultats de l'identification sur les maillages.

Notre algorithme a détecté 180 défauts réels. Il manquait donc 5 défauts. Lors de l'étape de détection, le nombre des faux positifs (fausses détections) a été considérable. Cela représentait 762 défauts. La plupart de ces faux positifs, 655, a été supprimée par l'algorithme d'identification. Le pourcentage des points classés comme défaut est de 7,7 % (1 150 373 points sur 14 774 118 au total) dont 17% (196 114 points) sont des faux positifs. Le critère choisi pour évaluer la performance de la classification est la mesure  $F_1$  qui est la moyenne harmonique de la précision et du rappel, deux variables calculées à partir des nombres de faux et vrais positifs. Le résultat d'identification a été finalement prometteur avec une mesure  $F_1$  moyenne de 0,84. L'algorithme donnait le meilleur résultat pour les branches avec une mesure  $F_1$  de 0,89. Le défaut le mieux classé ensuite était le broussin avec une mesure  $F_1$  de 0,73. Les mesures  $F_1$  sur les cicatrices de branche et les petits défauts étaient plus faibles et valaient 0,57 et 0,55 respectivement. Le mauvais résultat de l'identification des défauts de ces deux types était dû à des faux positifs. En particulier, 75 morceaux d'écorce ont été identifiés comme des cicatrices de branche. Sur les hêtres, ces faux positifs ont pu être liés à la maladie du *Nectria* qui a néanmoins un faible impact sur la qualité du bois (McCullough et al., 2005). Les résultats ont montré une certaine robustesse par rapport aux caractéristiques de l'espèce, telles que la rugosité de l'écorce, l'âge, le diamètre et la forme du tronc.



**Figure 3.2** Visualisation des résultats de l'identification de types des défauts sur les maillages du Chêne 411 (a) et du Hêtre 103 (b). ■ représente un défaut de type branche, ■ représente une cicatrice de branche, ■ représente un broussin et ■ représente un petit défaut.

### 3.4 Conclusions et perspectives

Ce chapitre a présenté une méthode basée sur le classifieur Random Forests pour identifier les défauts sur la surface de troncs à partir de données TLS. Les données en entrée de ce classifieur sont les défauts potentiels issus de la méthode de détection présentée dans le chapitre précédent. Ces défauts ont été classés en 4 classes ou comme écorce. L'expérimentation menée sur 185 défauts présents sur des arbres d'espèces différentes a montré que l'algorithme peut classer les défauts avec une mesure moyenne de  $F_1$  de 0,84. Bien que cette valeur de  $F_1$  ne soit pas excellente, le résultat reste prometteur et pourrait être amélioré de plusieurs manières. La première amélioration qu'on peut faire est d'entraîner le Random Forests par des données supplémentaires qui vont aider à augmenter le taux de reconnaissance. Ceci devrait notamment permettre de mieux classer les faux

positifs causés par des morceaux d'écorce. Les données supplémentaires permettront de mieux couvrir la grande variation de formes des défauts et aussi de l'écorce. De nouvelles classes pourront être ajoutées telles qu'une classe pour le *Nectria*. Une autre piste d'amélioration est d'utiliser un Random Forests pour chaque espèce d'arbres , mais cela nécessite plus de données que celles utilisées dans l'approche actuelle.

Les informations sur le type de défauts en plus de leurs dimensions et position peuvent être utilisées pour évaluer automatiquement la qualité du bois rond ou de l'arbre sur pied. Ceci est un premier pas vers le développement d'outils destinés à aider les experts dans l'évaluation de la qualité des arbres sur pied ou des grumes de bois en bord de route ou pour améliorer les informations provenant de scanners de forme dans l'industrie de première transformation du bois.

### **3.5 Article 4 : A machine-learning approach for classifying defects on tree trunk using terrestrial LiDAR (Projet d'article)**

# A machine-learning approach for classifying defects on tree trunk using terrestrial LiDAR

Van-Tho Nguyen<sup>1</sup>, Thiéry Constant<sup>1\*</sup>, Bertrand Kerautret<sup>2</sup>, Isabelle Debled-Rennesson<sup>2</sup> and Francis Colin<sup>1</sup>

<sup>1</sup> Université de Lorraine, AgroParisTech, Inra, Silva, F-54000 Nancy, France; [firstname.lastname@inra.fr](mailto:firstname.lastname@inra.fr)

<sup>2</sup> LORIA, UMR CNRS 7503, Université de Lorraine, 54506 Vandœuvre-lès-Nancy, France; [firstname.lastname@loria.fr](mailto:firstname.lastname@loria.fr)

\* Correspondence: [thiery.constant@inra.fr](mailto:thiery.constant@inra.fr); Tel.: +33-383-394-066

**Abstract:** The objective of this work is to investigate the potential for grading roundwood and standing tree from high density 3D point cloud recorded by TLS (Terrestrial Lidar Scanning). Among the attributes affecting the roundwood quality, our study is focused on the local defects due to the tree development, and particularly due to branching. These defects can significantly degrade the wood quality according to their number, type and size. In a previous work, we successfully developed the detection of the surface defects. The principal objective of this work is now to classify these detected defects into different types, such as branch, branch scar, epicormic shoot, burl and smaller defects. Our machine learning approach is based on a random forest using shape descriptors of defects, including Hu moment invariants, dimensions and species. The results of our experiments showed that most defects were detected and classified, even the small defects, with diameter as small as 0.5 cm. The defects were classified with an average  $F_1$  score of 0.835. The proposed method is robust to test tree species which are commonly commercialized in France, especially oak, beech, fir and pine.

**Keywords:** roundwood quality; random forest; standing tree grading

---

## 1. Introduction

Grading standing trees and roundwoods is an important task in a forest industry, during pre-harvest or in the sawmill [1]. The presence of defects indicates local variation of the material properties corresponding to a decrease of the normal distribution of the clearwood properties and characteristics and impacting the future products in their mechanical, physical or aesthetical functions. The grading of roundwood corresponds to a global assessment of the quality by taking many criteria into account through grading rules. Nevertheless, the grading result does not contain the precise information used by the process. The reference method to assess the quality of roundwood is to inspect the internal information recovered from X-ray tomography [2–5]. On the one hand, the X-ray tomography can achieve a very good accuracy for defect recognition (up to 95% [2]) or is able to detect the defects as small as 1mm in diameter by manually placing plot markers along the tracks of knot [5] or automatically detecting knots [6,7]. On the other hand, it has its own limitation, regarding the investment cost and the need to fell the tree and cut it down to logs. Recent studies confirmed a strong correlation between internal and external defects [8–11] with coefficients of determination ( $R^2$ ) greater than 0.6. This correlation suggests that the external defects can be used to estimate the quality of roundwoods or standing trees.

Lidar (light detection and ranging technique) can measure an object in 3D dimensions, using technique whereby laser beam is emitted and the reflected light is received by the detector. The data provided by Lidar is a point cloud which contains the three spatial dimensions (x, y and z coordinates) of the scanned object. In forestry, the Terrestrial Lidar Scanning (TLS) can provide information about an individual tree or a plot. A variety of forestry applications have been developed during the past two decades [12–21]. In particular, a number of studies take advantage of the potential of Lidar for the



replacement of conventional methods of measuring forest inventory parameters, such as tree height, diameter at breast height (DBH, trunk diameter measured at 1.3m above ground level), stand density, stand basal area and volume for biomass assessment [13–20,22].

On standing trees there have been attempts at estimating tree quality criteria from TLS [21,23], Airborne Lidar [21,24–26], or both types of Lidar [27]. The quality parameters targeted in these approaches are concerned mainly with the overall shape of the timber: ovality, curvature, and taper including the presence of branches. There is, a scarce number of research focused on the detection of external defects [9,28–30]. These research were mostly aimed at detect the large and very obvious defects. Thomas *et al.* [29,31] detected, on red oak and yellow poplar, defects with a diameter greater than 7.5 cm and protruding at least 2.2 cm from the bark. Kretschmer *et al.* [30] proposed an approach to detect and manually measure the branch scars on scots pine by highlighting them on a 3D reconstruction of the bark surface. The bark surface is colored based on the distance to a fitted cylinder surface corresponding to a trunk part. The scars, with a diameter of at least 2 cm and protruding at least 1.5 cm from the bark, were detected. The existing research about the automated classification of defect on tree bark using TLS are even scarcer. Schütt *et al.* [28] presented a semi-automatic approach, based on a neural network, to detect and classify wood defects using both range and intensity information of TLS data. However, neither information about how to train the neural network nor the result was presented. In a previous work [32], we successfully developed the detection of the surface defects which can detect defects with a diameter as small as 1 cm with a suitable spatial resolution of the 3D data. This important improvement was obtained thanks to the definition of the trunk centerline and a computation of the reference distance to the centerline for each individual point by taking its neighborhood into account. The computation of reference distance for each individual point allows to more precisely detect the abnormalities on the bark than more global reference surface based on primitive fitting such as cylinder [9,28,30] and circle [29,31]. Once the defects are detected, an automatic procedure must be able to recognize their types. The main challenge in the classification of these defects is the variability of their appearance, even for the same type of defect. In the forestry domain, the defects are often defined by the biological origin [33] that leads to a high intra-class variability and inter-class similarity. Figure 1 (a)-(d) and (f)-(h) give examples of the intra-class variability between branch scars and burls respectively. Inter-class similarity between a burl and an epicormic shoot is showed on Figure 1 (f) and Figure 1 (i). Factors contributing to the intra-class variability or inter-class similarity are the tree species, characteristics of its bark, the shape and the age of the defect and all the history of its development in connection with the environment of the tree.

There are several methods in the field of pattern recognition that can be applied to classify objects, such as neural networks [34], support vector machines [35], Random Forests [36], Bayes classifier [37], deformable models [38]. Most approaches are based either on parametric models or on machine learning techniques. In the remote sensing domain, the machine learning supervised classifiers are widely used since they are more flexible in handling the high variability in object appearance and are more robust than model-based approaches [39]. In particular, the Random Forests is a supervised machine learning method which is based on ensembles of classification trees. Random Forests exhibits many interesting properties, such as high accuracy, robustness against over-fitting, noise or missing data in the training set [40]. Moreover, the Random Forests is a non-parametric method that does not require the information on the distribution of data. These advantages make Random Forests a very successful classification method since its introduction by Breiman [36]. In the domain of remote sensing the Random Forests were used in landcover classification or urban area classification from the airborne Lidar [41,42] or landsat data [43–45]. In the forestry domain, the Random Forests were used to accompany the forest inventory, such as for biomass assessments [46], using the airborne Lidar. Othmani *et al.* [47] used the Random Forests to identify the tree species from the analysis of tree bark pattern from the mesh derived from TLS data. The Random Forests was used to assess the timber quality of Scots Pine by estimating tree properties, such as stem diameters, tree height and branch heights using the parameters computed from TLS data [21].

In this context, the objective of this work is to classify the potential defects detected on trunk surface by previously developed algorithms [32]. The types of defects are branch, branch scar, burl and small defect, including spheroblaste, bud cluster and picot. The chosen types were defined to represent the existing diversity of defects, but for some they were grouped together, because of difficulty for distinguishing them considering their size or their shape. We aim to develop a generic method that works on the common commercial tree species, especially oak and beech which have very different bark roughness, defect types and shapes. The former species has a furrowed bark and its most common defect types are burl and picot. The latter has a smooth bark and the most common defect type is branch scar with an eye brow (or "Chinese mustache") shape. Other species were also tested with our algorithms, and were conifer, such as fir, pine and spruce with different bark structures.

## 2. Materials and Methods

### 2.1. Defects on trunk surface

Several defects on the trunk surfaces can be caused by exogenous factors depending on their environment, such as heat, frost, other trees, animals, and human beings. Here our study focused on the most frequent source of defects, which arise from tree branching. Branching defects are the result of the development and growth of the tree. Their scars are in general associated with more or less protruding regions that result from the inclusion of the defect by the radial growth of the trunk. More precise definitions of what we consider as a branching defect is as follows:

- A sequential branch is a branch that emerged after a winter long rest of the original bud.
- An epicormic branch is a branch that emerged after several winters from a latent bud.
- A branch scar is a track of a branch either sequential or epicormic that maintains when this branch has died and is degraded.
- A bud is a miniature leafy shoot protected by protecting scales.
- A burl is a group of juxtapositional defects of one or more types, such as bud, picot, branch or branch scar. By definition, a burl can have a great variability in shape and size and composition.
- A bud cluster is a limited group of buds of less than 6 buds.
- A spheroblaste is a bud whose base produces xylem that progressively covers the apical meristem of the bud (mainly on beech).
- A picot is a small branch with its apex naturally pruned.

Typical defects on trunk surface are represented in Figure 1. These defects are characterized by a large intra-class variability in size and shape. For instance, burls can range from a large bud cluster with at least 6 buds to a very extended mass of buds, picots, short or long branches with a diameter of several tens of cm.

### 2.2. Methodology

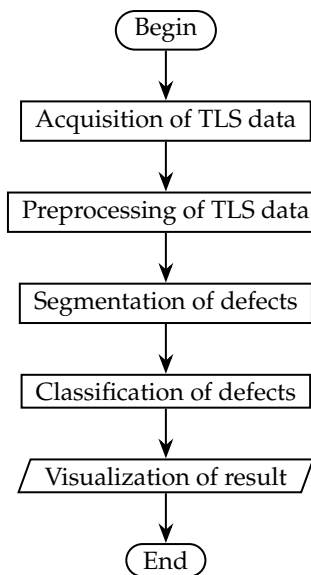
Our methodology is described as follows. The different steps are presented in Figure 2 and will be developed in details in the following sections. After their acquisition, the TLS data were preprocessed to obtain a smooth mesh corresponding to a trunk portion. Next, the defects were detected by using a segmentation algorithm, which is an improved version of the previously published work [32] and will be summarized in the section 2.5. Then, the detected defects were classified using a trained Random Forests. Finally, a visualization of the results was provided by various colors on the mesh according to the defect type. The classification was validated by comparing the results with the ground truth labels classified by an expert on the trees before the TLS scans were carried out.

### 2.3. Acquisition of TLS data

The tree stems exhibiting different defects were measured with the scanner Faro Focus 3D X130 in the Champenoux and Haye forests in the Lorraine region of France. In order to detect the small



**Figure 1.** Some illustrations of the types of defects considered in this study.



**Figure 2.** Overview of the processing flow for classifying the surface defects onto a trunk.

defects, we utilized a high resolution setting and put the scanner close to the trunk. The utilized resolution was 1/2 of the maximum value and the distance from scanner to trunk was approximately from 3 m to 4 m. With this setting, the resolution of the scan is 20,480 in 360° in both horizontal and vertical directions, the distance between to points in the point cloud was around 1 mm. The trees were sampled according to several criteria. Among the main commercial species, they must have a diameter

greater than 17 cm and represent a variability in bark roughness, which depends on the species and the age of trees. In regard to these criteria, we scanned 26 trees, including 9 European oaks (*Quercus petraea*), 8 European beeches (*Fagus sylvatica*), 3 wild cherries (*Prunus avium*), 2 pines (*Pinus sylvestris*), 3 firs (*Abies alba*) and 1 Norway spruce (*Picea abies*). These scans were divided into 2 sets. One was used to train the Random Forests and another was used to test the method efficiency (Table 1). Each defect was identified by a paper label with a scissored shape corresponding to its type, pinned in the close vicinity of the defect before the scanning. The objective was to maximize the number and type of defects per scan. Thus, trunks were either scanned entirely with 4 scans from different points of view or just one side described in 1 or 2 scans. If the trunk was scanned from multiple point of views, the scans were merged into a single file per tree to recover the 3D view of the trunk by the standard procedure, through the use of spheres and available in the FaroScene software.

**Table 1.** Number and attributes of the sample trees

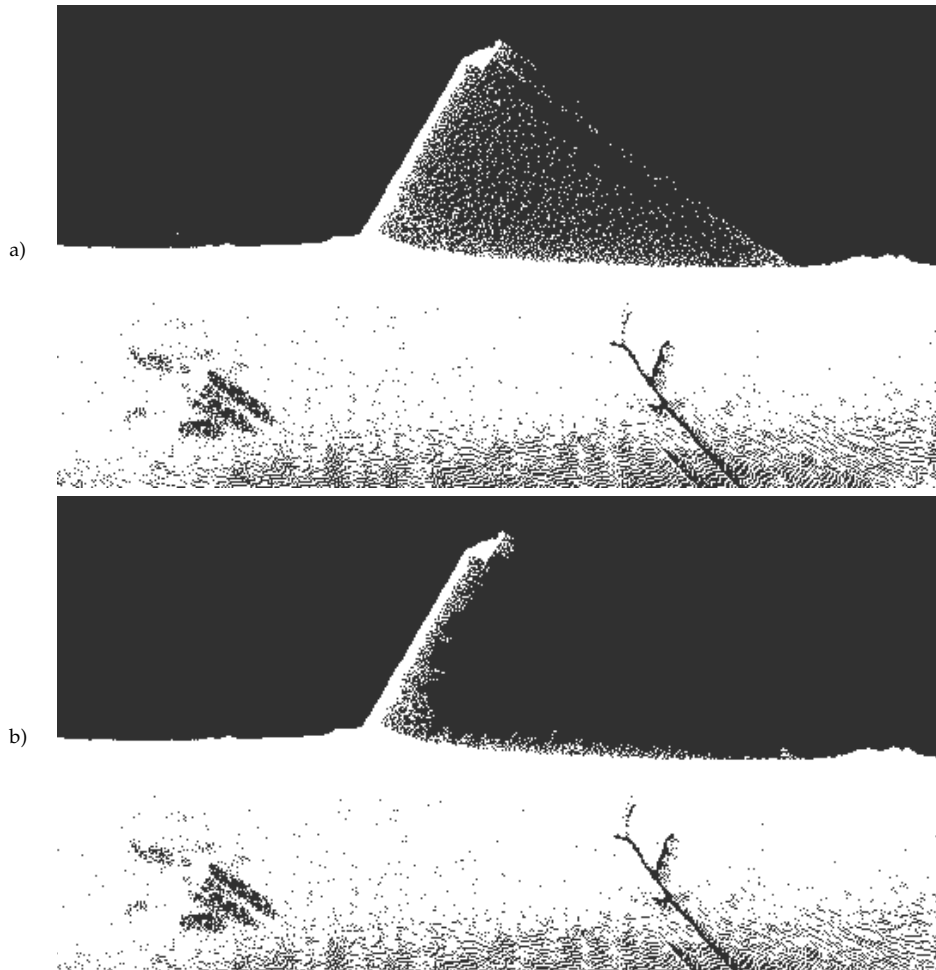
Species	Nb of trees		Range of diameters at breast height (cm)
	Train	Test	
Oak	6	3	35 - 76
Beech	5	3	30 - 57
Wild cherry	2	1	22 - 33
Pine	1	1	34 - 57
Fir	2	1	23 - 45
Spruce	0	1	19
Total	16	10	

#### 2.4. Preprocessing of TLS data

Lidar data are generally noisy, and the first processing step aimed to manage noise for enhancing the recognition rate. It included the reduction of noise and the smoothening of the trunk surface. The noise reduction is a difficult and complex process, due to different noise pattern from scan to scan. It depends on the condition of the environment of the scan and also the characteristics of the trees. For example, we observed that when the trunk has branches or small epicormic shoots, there was a lot of noise caused by the multiple interceptions of the same beam by several branches and the bark. This is the situation when a laser beam hits both the contour of the branch and the bark results in a phantom point in the middle of the branch and the bark (Figure 3 (a)). From our observation, the point density in noisy regions is often lower than in the relevant data regions. Thus, we proposed a simple approach to remove noise by clustering the point cloud by Euclidean distance with the idea that relevant data points are in the largest cluster where the point density is highest. The choice of the threshold is critical, if the threshold is too small, there is a risk that the relevant data would be removed, especially in the high part of the trunk where the resolution is lower. In our experiments, the best value for the threshold was 5 mm.

Due to the nature of a laser scan, the raw point cloud contains a certain level of error. For example the utilized scanner has a ranging error of  $\pm 2$  mm at 10 m. We observed that a smoothing step often produced a surface closer to the real surface. The following steps were performed for smoothing and creating a mesh from the trunk point cloud using the Graphite software (<http://alice.loria.fr/index.php/software.html>):

1. Smooth the point cloud with the used default parameters, which were 1 iteration and 30 neighbors.
2. Reconstruct the trunk surface by using the meshing procedure with the default parameter, which was radius of 5 mm.
3. Smooth the created mesh by using the remesh smooth function, the used parameter was the number of points similar to the one of the original point cloud.



**Figure 3.** Noise processing for a wild cherry trunk. Point cloud before (a) and after (b) noise reduction process that only keeps the biggest cluster; the minimal distance between two clusters is 5 mm.

### 2.5. Segmentation of defects

As presented in Figure 2, our strategy to classify the defects on trunk surface was first to detect them by a segmentation algorithm. The algorithm is an enhanced version of our previously published one [32] that focuses on defects with little protuberance from tree bark. In this study we proposed an additional step to segment these branches.

The motivation for this approach is that when there is a defect on the bark surface, there is a modification of the bark, typically a protruding region. To detect these regions, we use the centerline of the trunk as a reference. In the evaluation of the algorithm presented in [32], the presence of branches was identified as an inconvenience for detecting smaller defects in the vicinity. Thus, in the current version, the branches were first segmented by an algorithm which separates the points into two disjointed sets: (1). The set  $T$  contains the closer points to the trunk surface, and (2). The set  $B$  contains the branches according to the following algorithm:

- Estimation of the trunk radius  $r_m$ , using the mode of the distance to the centerline of all points in the point cloud.
- Division of the point cloud volume into slices with a thickness of  $l$  mm, following the centerline direction. Each slice is then divided into angular sectors with an angle of  $\frac{l}{r_m}$  radian. In our experiment, the best values for the  $l$  parameter was between 50 mm and 100 mm.

- For each angular sector, the nearest point to the center of the trunk is added to the set  $T$ , the other points of the portion are added to the set  $B$ .
- For each point  $P$  in the set  $T$ , we find the subset  $S$  of the set  $B$ , such as the distance between a point  $S_i \in S$  and  $P$  is less than or equal to  $\sqrt{2}l$  and move them in the set  $T$ .

After the branches are segmented, the original method proposed in [32] is applied to the set  $T$  as follows. (i) For each point  $P$  in the set  $T$ , we estimate a reference point  $\hat{P}$  from a linear regression linking the radius variation to longitudinal position on a patch of neighbor points of  $P$ . (ii) The defect points are detected by thresholding the difference between the distance from  $P$  and  $\hat{P}$  denoted ( $\delta$ ). (iii) The threshold is automatically computed on the histogram of  $\delta$  by the Rosin's method [48]. The detected defect points are then merged with the set  $B$  containing the branches to form a set of defect points  $D$ . The different potential defects were obtained by clustering the defect points  $D$  by using Euclidean distance.

## 2.6. Classification of defects

### 2.6.1. The random forest classifier

Random Forests [36] classifier is an ensemble classifier that aggregates a set of CARTs [49] (Classification and Regression Trees) to make a prediction. In the training step, all trees are built with the same parameters but on different subsets of the training samples. These subsets are generated from the training samples by a bootstrap sampling, which randomly selects the same number of vectors from the original set and the remaining "out-of-bag" (OOB) is used to compute the estimation error, which is known as the OOB error. Unlike CART, the Random Forests classifier does not consider all variables at each node to determine the best split threshold but a random subset of variables of the feature vector and the trees are built without pruning. The cardinality of the subset is an input parameter. Another important parameter of Random Forests classifier is the number of trees, which must be sufficiently large to capture the full variability of the training data and yield good classification accuracy. One of the advantages of Random Forests classifier is that it does not overfit when increasing the number of the trees at the expense of a slower running time. In the classification step, Random Forests classifier tests the feature vector, describing the new object with each tree in the forest. Each tree makes a classification, or in other words, gives a vote for a class. The random forest chooses the class that is voted by the majority of trees.

As mentioned above, the number of trees in the forest ( $nbTrees$ ) and the number of variables ( $nbVariables$ ) used to select and test for the best split when growing the trees are two important input parameters needed to train Random Forests classifier. The OOB error can be used to find the optimal value for these parameters. We ran an experiment with the  $nbTrees$  from 100 to 5000 and the  $nbVariables$  from 1 to the number of variables of the feature vector. For each value of  $nbVariables$ , we can find the minimum value of  $nbTrees$ , which gives the minimum OOB error. The computing time required to train the Random Forests is:

$$T\sqrt{MN\log N} \quad (1)$$

where  $T$  is  $nbTrees$ ,  $M$  is  $nbVariables$ , and  $N$  is the number of training samples [36].

The Random Forests had been implemented in a number of free and open source libraries. In this study we used the implementation in OpenCV [50]. The advantage of OpenCV is that it is compatible with the implementation of our algorithms in C++ programming language.

### 2.6.2. Feature vector

In this step, we take the defects detected by our segmentation algorithm and construct the feature vector based on our expertise on the defects. Before computing the feature, the point cloud of defect is converted from a Cartesian coordinate system to a custom coordinate system  $\{l, z, d\}$  where  $l$  is

the arc length,  $z$  is the height and  $d$  is the difference between the distance and the reference distance from the point to the centerline (the distance between  $P$  and  $\hat{P}$  as presented in the section 2.5). This conversion allows us to measure the defect diameter along the curve surface of the trunk similar to a manual measurement. To reduce the affect of the inhomogeneity of point clouds due to the scan settings, the feature vector is computed from a subsampled point cloud. The subsampled point cloud is computed by keeping only the closest point to each voxel of a regular voxel grid of the defect point cloud. The voxel size is chosen by the average point spacing, which is 3 mm in our experimentation. The following features are used:

1. Species:  $s$ .
2. Ratio between the number of points of the defect and the volume of its bounding:  $c$  (equation 2)
3. Defect arc length:  $w = l_{max} - l_{min}$ .
4. Ratio between  $w$  and defect height:  $\frac{w}{h}$  where  $h$  equals  $z_{max} - z_{min}$ .
5. Ratio between  $w$  and maximum of  $d$ :  $\frac{w}{d_{max}}$ .
6. Mean of difference between the distance from  $P$  and  $\hat{P}$  for all points  $P$  of the defect:  $\bar{d}$ .
7. Standard deviation of the difference between the distance from  $P$  and  $\hat{P}$  for all points  $P$  of the defect:  $\sigma_d$ .
8. Hu moment invariants:  $I_1, I_2, I_3, I_4, I_5, I_6, I_7$  (see equation from 5 to 11).
9. Ratio between the eigenvalue  $\lambda_1$  and the eigenvalue  $\lambda_3$ :  $\frac{\lambda_1}{\lambda_3}$ .
10. Ratio between the eigenvalue  $\lambda_2$  and the eigenvalue  $\lambda_3$ :  $\frac{\lambda_2}{\lambda_3}$ .
11. Angle between the eigenvector  $\vec{v}_3$  and the trunk axis at the height of defect:  $\alpha$ .

where  $\lambda_1$  is the eigenvalue associated with the eigenvector  $\vec{v}_1$  of the defect having the smallest angle, with the radial vector of the trunk at the center of the intersection between the defect and the trunk.  $\lambda_2$  is the eigenvalue associated with the eigenvector  $\vec{v}_2$  of the defect having the smallest angle with the tangential vector of the trunk at the center of the intersection between the defect and the trunk.  $\lambda_3$  is the eigenvalue associated with the eigenvector  $\vec{v}_3$  of the defect having the smallest angle with the trunk axis at the height of defect.

The species is an important variable as each species has a specific set of defects. For example, oak has burls but does not have spheroblaste, that is conversely related to beech. In addition, for the same defect type, its shape can be different from one species to another. For example, a branch scar on oak and on beech is very different.

Another relevant variable is the ratio between the number of points of the defect and the volume of its bounding box, which measure the compactness of the defect in the  $\{l, z, d\}$  coordinate system equation 2. This feature can discriminate a flat defect and a significantly protruding defect.

$$c = \frac{\text{number of points}}{(l_{max} - l_{min})(z_{max} - z_{min})(d_{max} - d_{min})} \quad (2)$$

By using our expertise in the domain, the classification of defects, in particular the small defects, such as burl and cluster of bud, the dimension of defect is an important criterion. For that reason, we include the arc length  $w$  as a feature. The ratio  $\frac{w}{h}$  allows us to distinguish between a branch scar and a bark portion on oak tree since the branch scars often have width greater than height and bark portions have width smaller than height. The ratio  $\frac{w}{d_{max}}$  helps to distinguish a flat object, such as bark portion, branch scar and a more protruding one such as spheroblaste and picot. The mean and standard deviation of  $d$  were also included in the feature vector because they can aid to distinguish between a branch scar and a burl composed only in buds.

The Hu moment invariants [51] have good characteristics for the object recognition since they are invariant with respect to translation, scale, and rotation. The Hu moment invariants  $\{I_1, \dots, I_7\}$  are computed from the normalized central moments  $mu_{ij}$  (see equation from 4 to 11).

$$mu_{ij} = \sum_{z,l} (z - \bar{z})^i (l - \bar{l})^j d \quad (3)$$

$$nu_{ij} = \frac{mu_{ij}}{mu_{00}^{(i+j)/2+1}} \quad (4)$$

$$I_1 = nu_{20} + nu_{02} \quad (5)$$

$$I_2 = (nu_{20} - nu_{02})^2 + 4nu_{11}^2 \quad (6)$$

$$I_3 = (nu_{30} - 3nu_{12})^2 + (3nu_{21} - nu_{03})^2 \quad (7)$$

$$I_4 = (nu_{30} + nu_{12})^2 + (nu_{21} + nu_{03})^2 \quad (8)$$

$$I_5 = (nu_{30} - 3nu_{12})(nu_{30} + nu_{12})[(nu_{30} + nu_{12})^2 - 3(nu_{21} + nu_{03})^2] + (3nu_{21} - nu_{03})(nu_{21} + nu_{03})[3(nu_{30} + nu_{12})^2 - (nu_{21} + nu_{03})^2] \quad (9)$$

$$I_6 = (nu_{20} - nu_{02})[(nu_{30} + nu_{12})^2 - (nu_{21} + nu_{03})^2] + 4nu_{11}(nu_{30} + nu_{12})(nu_{21} + nu_{03}) \quad (10)$$

$$I_7 = (3nu_{21} - nu_{03})(nu_{30} + nu_{12})[(nu_{30} + nu_{12})^2 - 3(nu_{21} + nu_{03})^2] + (3nu_{12} - nu_{30})(nu_{21} + nu_{03})[3(nu_{30} + nu_{12})^2 - (nu_{21} + nu_{03})^2] \quad (11)$$

The eigenvectors and eigenvalues of the defect were computed from a principal component analysis (PCA) [52], which can be useful for distinguishing between the defect with a long axis (branch) and the flatter ones. Although we have yet to classify the sequential branches and epicormic branches, the angle between the eigenvector  $\vec{v}_1$  of defect and the stem axis can be used to classify these types of branch on beech, oak and fir, as epicormic branches are quasi perpendicular to the stem axis while sequential branches are more fastigiated.

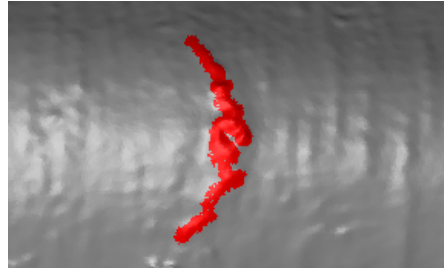
### 2.6.3. Construction of the training dataset

We used both manually segmented and automatically segmented defects to train the Random Forests. The manual segmentation was done by using a home-made software, based on the library DGTal [53]. The software allowed us to select the faces on the mesh to highlight the defects (Figure 4). Each segmented defect is then saved in a separate file and used for training the Random Forests. We also trained the Random Forests using the results of our segmentation algorithm, along with the verification given by the shape of paper labels marked next to defects prior to scanning trees. The Bark class was introduced even though it is not a defect type; it is often miss detected as a defect by the segmentation algorithm. This aligns with our approach as the detection step was built to provide all potential zones of defects assuming the risk of false positive that could be eliminated in the classifications step. The classes in our training database include:

1. Branch, including sequential branch and epicormic branch.
2. Branch scar.
3. Burl.
4. Small defects, including picot, spheroblaste, bud and bud cluster.
5. Bark.

Table 2 summarizes the number of defects of each type in the training set. As we can see in this table, the defects were not equally distributed among the species. The number of defects of each type was also not equal. The number of defects of each type depends on the variability of its shape. As expected, the burl and the branch scar were more numerous because they have a more variability in shape. In contrast, the number of branches in the training set was less numerous because their shape was relatively easier to recognize such as a very large protruding from the trunk. The data in bark class came from trees with furrowed bark, such as oak and fir.





**Figure 4.** Manual segmentation of a defect

**Table 2.** Summary of the defects and barks encountered in the training set.

Species	Branch	Branch scar	Burl	Small defects	Bark
Oak	7	3	159	63	116
Beech	34	51	26	20	2
Wild Cherry	15	5	0	0	0
Pine	0	4	0	0	0
Fir	0	38	0	0	10
Total	56	101	185	83	128

#### 2.6.4. Performance evaluation criteria

To evaluate the performance of the classification algorithm, we used the F-measure, which is one of the most commonly used performance measurements for classification problems. The F-measure is the harmonic mean of precision ( $P$ ) and recall ( $R$ ). We used the  $F_1$  score which has equal weights on  $P$  and  $R$ . The precision  $P$  is the number of correctly classified positive examples divided by the number of examples labeled by the system as positive (equation 12). The recall is the number of correctly classified positive examples divided by the number of positive examples in the data (equation 13). On a binary classification problem the  $F_1$  is defined by the equation 14.

$$P = \frac{TP}{TP + FP} \quad (12)$$

$$R = \frac{TP}{TP + FN} \quad (13)$$

where  $TP, FP, FN$  are true positive, false positive and false negative respectively. Their definition is as follows:

- $TP$  is the number of clusters correctly classified as defect.
- $FP$  is the number of clusters incorrectly classified as defect.
- $FN$  is the number of clusters incorrectly classified as non-defect.

$$F_1 = 2 \frac{PR}{P + R} \quad (14)$$

For a multi-classes classification problem, the F-measure must be extended from the binary classification by an average of the F-measure of each class. There are two approaches [54]. One approach is the macro-average F-measure, which is the unweighted mean of F-measure for each label. The other is the micro-average F-measure, which considers predictions from all instances pooled together and calculates the F-measure across all labels. In our experimentation, the database

is unbalanced, therefore the micro-average is better for measuring the overall performance. The micro-average F-measure  $F_{\mu 1}$  is computed as follows:

$$F_{\mu 1} = 2 \frac{P_{\mu} R_{\mu}}{P_{\mu} + R_{\mu}} \quad (15)$$

$$P_{\mu} = \frac{\sum_{i=1}^n TP_i}{\sum_{i=1}^n (TP_i + FP_i)} \quad (16)$$

$$R_{\mu} = \frac{\sum_{i=1}^n TP_i}{\sum_{i=1}^n (TP_i + FN_i)} \quad (17)$$

where  $n$  is the number of classes.

We also used the confusion matrix [55] to evaluate the performance for a more detailed analysis of the misclassification between classes.

### 3. Results

In this section, we present the results of the segmentation algorithm, followed by the results of the classification of these detected defects. First, we compare the results of segmentation algorithm with the classification algorithm using the criterion of defect detection. In this analysis, we want to assess the detected defects with the observed defects regardless of the defect type, since in the evaluation process of wood quality, the number of defects is considered one of the most important indicators. Second, we want to assess the performance of the classification algorithm by defect type. Last, we tested the efficiency of our algorithm in regard to tree species and other parameters.

#### 3.1. Defect detection

Table 3 shows the results of the segmentation and classification algorithms for each individual tree in the test data base in term of defect detection. We can see that the segmentation algorithm detected almost all the defects, with 180 detected of 185 in total. However, the number of false positives was very high (762), as in this step, we aimed to detect as many defects as possible and accepted the false positives, including the bark portions, which will then be removed by the classification algorithm. Table 3 also shows that these false positives were mostly removed by the classification algorithm at the expense of some defects lost. The classification algorithm removed 655 (86%) false positives but also 17 (9.4%) correctly detected defects. This led to a higher precision rate and a slightly lower recall rate, when compared to the corresponding criteria of the segmentation algorithm. Thus, the  $F_1$  score of the detection result of the classification (0.72) was hugely improved compared to the one of segmentation which is 0.32 (see Figure 5). We also observed that the segmentation algorithm produced more false positives on trees with furrowed barks, such as oak and pine than on trees with smooth barks, such as beech and wild cherry. By contrast, the classification algorithm removes the false positives more efficiently on trees with furrowed bark than on trees with smooth-bark. For example, in Table 3 we can see that on pine the number of false positives from the segmentation and classification are 103 and 4 respectively while on Beech 102 these numbers are 70 and 22. The difference is illustrated in Figure 6.

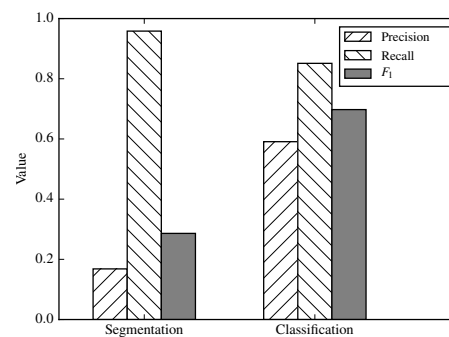
#### 3.2. Defect classification

Figure 7 presents examples of the classification results by coloring the mesh in agreement with the defect type.

Figure 8 shows the result of the classification of defects. The overall  $F_{\mu 1}$  score is 0.84. The algorithm did not perform equally well on all classes of defect. The branch had the best  $F_1$  score of 0.89, followed by the burl with  $F_1$  score of 0.73. The algorithm performed less well on branch scar and the small defect types with  $F_1$  scores of 0.57 and 0.55 respectively. Therefore, we will use the confusion matrix to explain these results.

**Table 3.** Results of the segmentation (seg.) and classification (cla.) steps compared with the observed defects.

Tree name	Observed	Correct detected		False Positive		Missed	
		Seg.	Cla.	Seg.	Cla.	Seg.	Cla.
Oak 334	8	8	8	9	0	0	0
Oak 411	25	24	19	147	7	1	6
Oak 569	24	23	20	79	7	1	4
Beech 101	30	30	24	55	19	1	6
Beech 102	29	25	21	63	12	0	8
Beech 104	24	22	18	47	14	2	6
Wild cherry	8	8	8	9	2	0	0
Pine	4	4	4	103	4	0	0
Fir	14	14	14	129	8	0	0
Spruce	18	18	16	111	8	0	2
Total	185	180	163	762	107	5	22



**Figure 5.** Precision, recall and  $F_1$  score of the defect detection resulting from the segmentation and classification algorithms.

We observed that the false positives did not equally occur on all tree species. In particular, the number of false positives on beech samples was greater than on oak samples (Table 3).

Figure 9 represents the confusion matrix of the classification result. The matrix shows the matching between the classes. From the confusion matrix, we can see that one branch was classified as branch scar. A more detailed analysis showed that it was a dead branch of a wild cherry, with a very short branch (the large green region in Figure 7 (c)). Another branch was classified as a burl because a burl with an epicormic branch may have an appearance of an epicormic branch.

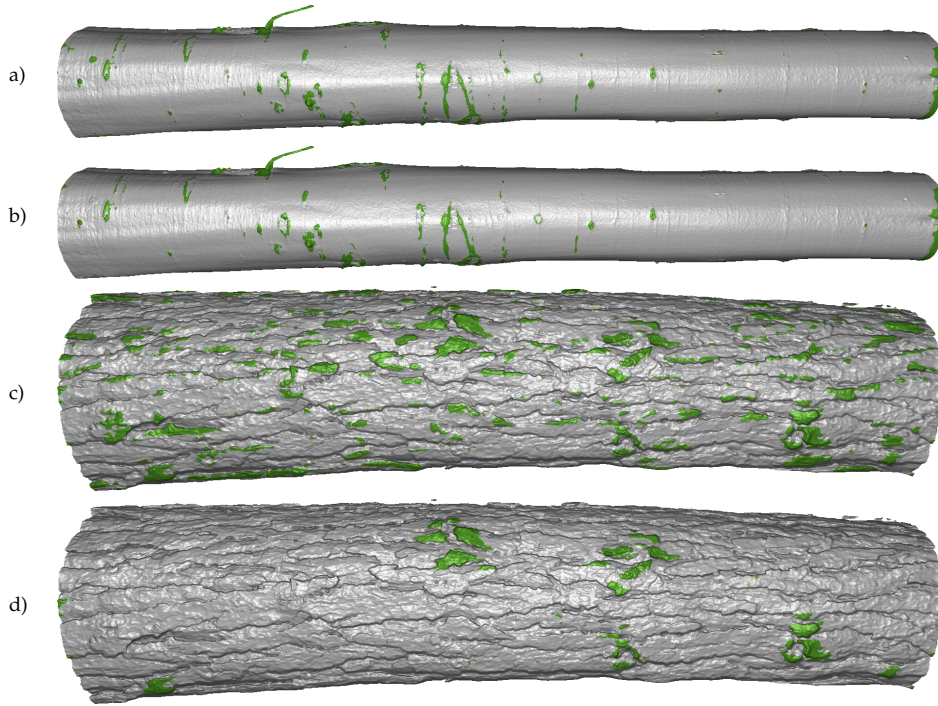
While the recall of algorithm on the branch scar was very high (0.89), the precision was not as good (0.46) because there were 75 bark portions recognized as branch while there were 101 branch scars in total. The confusion matrix shows that the small defects were often confounded with bark, branch scar and burl. For an illustration of the similitude between a small defect and a burl, see Figure 1 (h) and (j). It is to be noted that the number of bark portions miss classified as small defects was 20.

Although no spruce data were used to train the Random Forests the classification result on this spruce (Figure 7 (d)) was good, as 15 defects out of 18 were detected. However, 2 branch scars were detected as small defects and 2 branch scars were detected as burls.

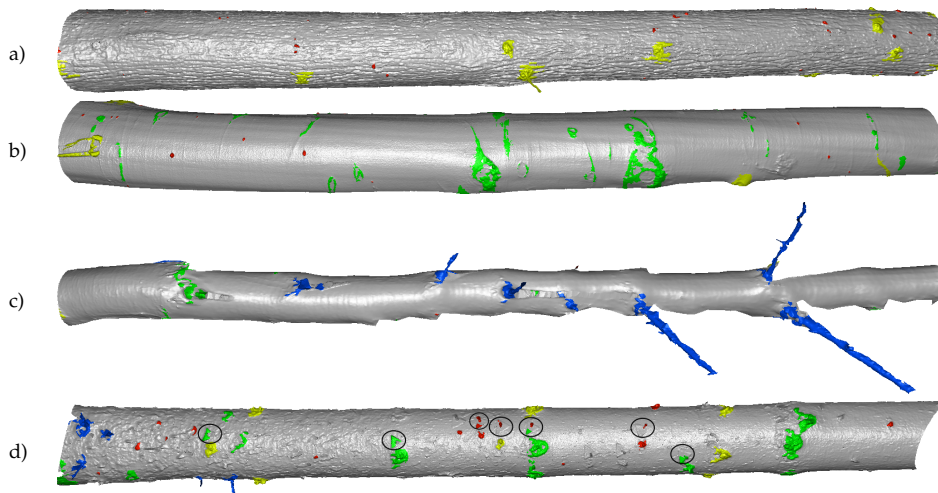
## 4. Discussion

### 4.1. Defect detection

Our algorithms had a very good performance in defect detection, even with the small defects. That was because we have a good and robust estimation of the trunk centerline, allowing the calculation of a good reference distance to detect even a slight modification of bark roughness. It is clear that there

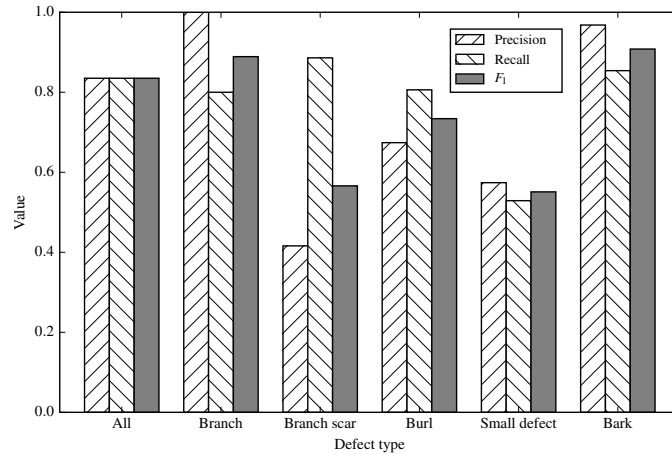


**Figure 6.** Defects detected by the segmentation algorithm (a,c) and refinement by the classification algorithm (b,d) for two logs: Beech 102 (a,b) and Pine (c,d).

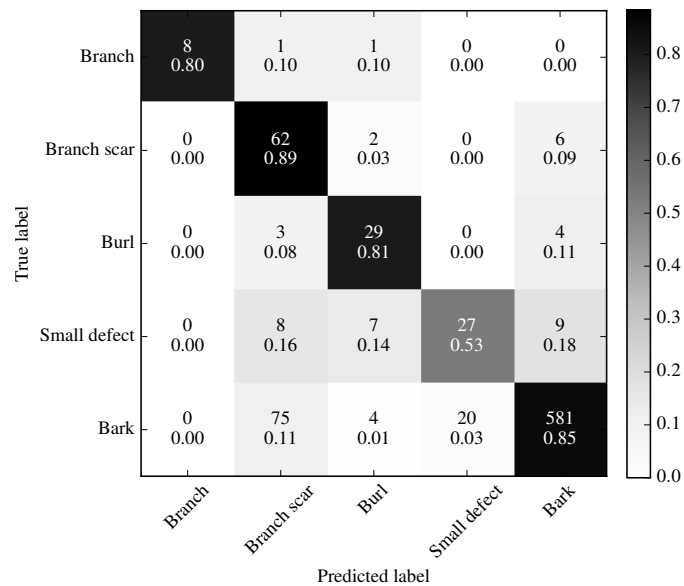


**Figure 7.** Examples of the classification results on the mesh of the Oak 569 (a), Beech 101 (b), Wild cherry 3 (c), and Spruce (d). **Blue** is Branch type including both sequential and epicormic branches, **Green** is Branch scar, **Yellow** is Burl, **Red** is small defect including bud cluster, spheroblaste and picot. On the mesh of the Spruce, we surrounded the paper marks and pushpins that were classified as defects. These false positives were ignored in our evaluation.

are several factors which can impact the detection, such as scan resolution and quality, missing data resulting from occlusion. Our algorithm can detect defects with a 0.5 cm diameter, thus outperforming all previous works with size ranging from 7.5 cm [29] to 2.0 cm [30]. Moreover, Kretschmer *et al.* [30] was only focused on the branch scars and their method was not automatic.



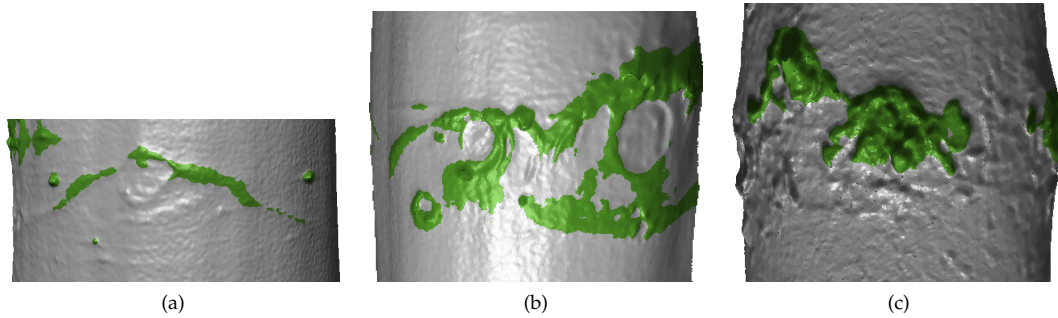
**Figure 8.** Precision, recall and  $F_1$  score of the different defect types



**Figure 9.** Confusion matrix with the absolute value and normalized value, the color of cells is a function of normalized value

Another important point related to the defect detection is the cross-sectional shape. Stängle *et al.* [9], Thomas *et al.* [29], Kretschmer *et al.* [30] used an estimation of cylinder or circle to create a reference surface which implied that when the cross section of a tree is not perfectly round, the performance of the algorithms might be degraded. Unlike these approaches, the performance of our method was based on narrow longitudinal patch suffers a minimal impact due to the shape of trunk cross section [32]. Moreover, the results showed that our method can work with partial data, as most of our samples were only on one side of trunks.

On beech and wild cherry trees the branch scars often have a "Chinese mustache" shape (or eye brow). It is often large and the two parts of the mustache are often thin. This may result in an over-segmentation (Figure 10 (a)). Two or more large defects can also be close enough to form a large shape. Although we considered the large shape as multiple separated defects, the algorithm saw it as a defect which consequently created an under-segmentation (Figure 10 (b)). The under-segmentation can also occur on the trunk of conifer, in the case of the connected branch scars (Figure 10 (c)).



**Figure 10.** Example of over-segmentation on Beech (a), under-segmentation on Beech (b) and on Spruce (c)

As the segmentation was the preceding step of the classification, the performance of the classification algorithm depended on the performance of segmentation step. Thus, any improvement in segmentation will result in a better classification. Since the segmentation algorithm had several parameters, the choice of these parameters had an impact on the final results. The most important parameters were the patch size and the bin width of the histogram used to find the threshold by the Rosin's method [48], and the voxel size used to compute the centerline, which are all discussed in details in [56]. As mentioned earlier, the over and under-segmentation can occur in the segmentation step. These errors can affect the classification, and consequently, the assessment of the tree quality. In addition to the influence in miss-classification, the over-segmentation increased the number of defects and decreased their dimension. By contrast, the under-segmentation decreased the number of defects and increased their dimensions.

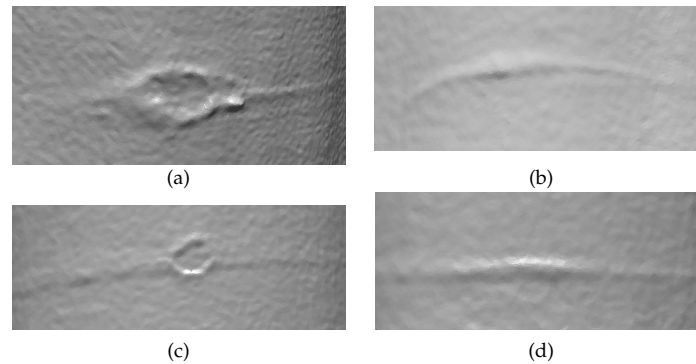
#### 4.2. Defect classification

Visually, we can see that our algorithms were able to detect and classify most of the defects (Figure 7), even the small defects. However, we did not obtain a very high  $F_1$  score, because first, we could not totally remove all the false positives and second, there was some confusion between the classes due to the very high intra-class variability. This variability combined with the inter-class similarity led to a high confusion. For example, the confusion between branch scars and burls can be explained by the fact that some burls containing only buds have a shape looking like a branch scar, as both are flat. On the field, the human eyes can easily distinguish these two defect types, however, in the point cloud or mesh, it could be difficult to distinguish them. The confusion between burls and small defects can merely be explained by the definition of burl and small defect. In our database, small defect types include several biological defect types: bud cluster with less than 6 buds, spheroblaste and picot of which a bud cluster may have a shape similar to a small burl. The confusion was high, even for expert eyes. Finally, we observed that a large amount of confusion occurred due to the botanic definition of defects. In particular, the burl has a very large range in dimension ranging from 2 cm to several tens of cm. The shape of a burl also varied largely depending on its composition. It can be quite flat when it is composed only of buds or look like a branch when it is composed of an epicormic branch and buds.

Moreover, our training database was not big enough to cover all the variability of defects, in particular branch scars and small defects. This suggests that more data is needed to increase the performance. We observed that in order to build the database by manually segmenting the defects is very time consuming. We can nevertheless accelerate this process by using the result of our program to feed the database with a verification by an expert.

#### 4.2.1. Influence of species and bark roughness

The worst classification performance observed on beech compared to on oak contradicts our hypothesis that, "the detection of defect on smooth bark is more efficient than on furrowed bark". Our experiments showed that our algorithms detected the defects on both variants equally (Table 3). In the segmentation step, on trees with furrowed bark, there were many more false positives, resulting of the miss-classification of bark portions as defects. This is in agreement with our hypothesis. However, the false positives on trees with furrowed bark have a common shape created by the pattern of the rhytidome is easily detectable and removed by the classification by using the bark class. By contrast, on trees with smooth bark, the false positives were created by bark portions very similar to actual defects in terms of protrusion and spatial distribution. Moreover, on species with smooth bark, and particularly on beech, we also observed a lot of wrinkles or cambium alterations revealed by the elliptical shape (*Nectria* disease) of bark (Figure 11). These alterations were often miss-classified by our algorithm as branch scars. Thus, the classification algorithm has a higher performance on furrowed bark trees than on smooth bark trees.



**Figure 11.** Examples of equivalent bark appearances considered either to be a branch scar (a) and (b) or a non-defect (c) and (d). This has been determined according to our own biological expertise. This figure illustrates the difficulty in distinguish between defects and non-defects. The region in (c) while having a shape similar to a branch scar as in (a), is a scar of slight damages due to *Nectria* attack affecting just the bark and not the wood below. The region in (d) has been considered as non-defect, since it was formed by the covering of a dead bud at the very first years of tree development with no consequence on the wood quality.

#### 4.2.2. Parameters of Random Forests and future improvements

Our experiments also showed that the optimal parameters of Random Forests are easy to find. Random Forests have only two principle parameters. Their performance was slightly influenced by the number of trees if it was chosen sufficiently high (1,000 trees in our experimentation). We also noticed that Random Forests are very robust to over-fitting so the feature selection is less critical. We can use all the features in agreement with the finding of Puissant *et al.* [57]. Since Random Forests can give a good performance, even with a small training dataset [44], these authors also suggested that it also depended on the intra-class variability. For classes with a high variability, more training data is needed. In regard to trunk surface defects, the intra-class variability of burl and branch scar is very high. Thus, in future work, we would want to add more training data of these types. As the incident angle of laser beams changes with the different heights of stem, it is also important to have the defect data from different heights.

Another suggestion to improve the performance of the classification is to remove species in the feature vector and separately train a Random Forests for each species, considering that the information on a species is a prerequisite brought by an operator or by another identification step [47]. This

approach might have a better performance but requires more training data. With a small database size, we found that using a Random Forests for all species gives a better performance due to the lack of training data for some species.

#### 4.3. Limitation of the TLS technology in grading tree quality

As a common problem for the remote sensing technologies, the quality of TLS data can be limited by the occlusion, especially when there are branches on trunk or false positive created by moss or parasite like ivy. However, scanning the tree from multiple views can reduce the level of occlusion; the occlusion on the high part of the tree is very difficult to avoid. In this work, we used a very high resolution for the 3D data, and it is obvious that is currently higher than the settings used in forest plot measurements. Nevertheless, the method is efficient and could be calibrated for different conditions.

## 5. Conclusions

In this paper, we have presented a Random Forests-based classifier for identifying defects on trunk surface from TLS data. The potential defects were firstly detected by our segmentation algorithm. Each detected defect was classified into one of the four defect classes or bark using the Random Forests classifier. Our experiment showed that we can detect and classify defects as small as 0.5 cm in diameter. The overall  $F_{\mu 1}$  score of classification algorithm was 0.835. Even though the score was not very high, these preliminary results were very promising by indicating that we can improve the score with the addition of more data. When we acquire enough data we can train the Random Forests separately for each tree species.

The information about the defect type in addition to its dimension and position can be used to assess the quality of roundwood or standing tree. This is a first step towards developments for helping experts in the assessment of quality of standing trees or timber logs in forests or to enhance the knowledge coming from true shape scanners in the primary wood processing industry.

**Acknowledgments:** This work was supported by the French National Research Agency through the Laboratory of Excellence ARBRE (ANR-12- LABXARBRE-01) and by the Lorraine French Region. We would like to thank Florian Vast for providing technical support and for scanning and measuring the trees used in this research.

#### Author Contributions:

Van-Tho Nguyen designed and implemented the algorithms, conducted the statistical analysis, and wrote the manuscript. Thiéry Constant supervised the experiment, helped with algorithm design, and contributed to the writing of the manuscript. Francis Colin supervised the experiment, contributed to the acquisition of data and the writing of the manuscript. Bertrand Kerautret and Isabelle Debled-Rennesson assisted with the design of the algorithms.



1. Fonseca, M.A. *The measurement of roundwood: methodologies and conversion ratios*; CABI, 2005.
2. Li, P.L.P.; a.L. Abbott.; Schmoldt, D. Automated analysis of CT images for the inspection of hardwood logs. *Proceedings of International Conference on Neural Networks (ICNN'96)*. IEEE, 1996, Vol. 3, pp. 1744–1749.
3. Zhu, D.; Connors, R.W.; Schmoldt, D.L.; Araman, P.a. A prototype vision system for analyzing CT imagery of hardwood logs. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics* **1996**, *26*, 522–532.
4. Aguilera, C.; Sanchez, R.; Baradit, E. Detection of knots using X-ray tomographies and deformable contours with simulated annealing. *Wood Res.* **2008**, *53*, 57–66.
5. Colin, F.; Mothe, F.; Freyburger, C.; Morisset, J.B.; Leban, J.M.; Fontaine, F. Tracking rameal traces in sessile oak trunks with X-ray computer tomography: Biological bases, preliminary results and perspectives. *Trees - Structure and Function* **2010**, *24*, 953–967.
6. Krähenbühl, A.; Kerautret, B.; Debled-Rennesson, I.; Longuetaud, F.; Mothe, F. Knot detection in x-ray ct images of wood. *International Symposium on Visual Computing*. Springer, 2012, pp. 209–218.
7. Krähenbühl, A.; Roussel, J.R.; Kerautret, B.; Debled-Rennesson, I.; Mothe, F.; Longuetaud, F. Robust Knot Segmentation by Knot Pith Tracking in 3D Tangential Images. *International Conference on Computer Vision and Graphics*. Springer, 2016, pp. 581–593.
8. Thomas, R.E. Modeling the relationships among internal defect features and external appalachian hardwood log defect indicators. *Silva Fennica* **2009**, *43*, 447–456.
9. Stängle, S.M.; Brüchert, F.; Kretschmer, U.; Spiecker, H.; Sauter, U.H. Clear wood content in standing trees predicted from branch scar measurements with terrestrial LiDAR and verified with X-ray computed tomography 1. *Canadian Journal of Forest Research* **2013**, *44*, 145–153.
10. Racko, V. Verify the accuracy of estimation the model between dimensional characteristics of branch scar and the location of the knot in the beech trunk. *Annals of Warsaw Univ. of Life Sciences-SGGW. Forestry & Wood Technology* **2013**, *84*.
11. Pyörälä, J.; Kankare, V.; Vastaranta, M.; Rikala, J.; Holopainen, M.; Sipi, M.; Hyyppä, J.; Uusitalo, J. Comparison of terrestrial laser scanning and X-ray scanning in measuring Scots pine (*Pinus sylvestris* L.) branch structure. *Scandinavian Journal of Forest Research* **2017**, pp. 1–8.
12. Dassot, M.; Colin, A.; Santenoise, P.; Fournier, M.; Constant, T. Terrestrial laser scanning for measuring the solid wood volume, including branches, of adult standing trees in the forest environment. *Computers and Electronics in Agriculture* **2012**, *89*, 86–93.
13. Hopkinson, C.; Chasmer, L.; Young-Pow, C.; Treitz, P. Assessing forest metrics with a ground-based scanning lidar. *Canadian Journal of Forest Research* **2004**, *34*, 573–583.
14. Simonse, M.; Aschoff, T.; Spiecker, H.; Thies, M. Automatic determination of forest inventory parameters using terrestrial laserscanning. *Proceedings of the ScandLaser Scientific Workshop on Airborne Laser Scanning of Forests*, 2003, Vol. 2003, pp. 252–258.
15. Popescu, S.C.; Wynne, R.H.; Nelson, R.F. Estimating plot-level tree heights with lidar: local filtering with a canopy-height based variable window size. *Computers and Electronics in Agriculture* **2002**, *37*, 71–95.
16. Thies, M.; Spiecker, H. Evaluation and future prospects of terrestrial laser scanning for standardized forest inventories. *Forest* **2004**, *2*, 1.
17. Henning, J.G.; Radtke, P.J. Detailed stem measurements of standing trees from ground-based scanning lidar. *Forest Science* **2006**, *52*, 67–80.
18. Litkey, P.; Liang, X.; Kaartinen, H.; Hyyppä, J.; Kukko, A.; Holopainen, M. Single-scan TLS methods for forest parameter retrieval. *Proceedings of SilviLaser* **2008**, *2008*, 8th.
19. Yao, T.; Yang, X.; Zhao, F.; Wang, Z.; Zhang, Q.; Jupp, D.; Lovell, J.; Culvenor, D.; Newnham, G.; Ni-Meister, W.; others. Measuring forest structure and biomass in New England forest stands using Echidna ground-based lidar. *Remote sensing of Environment* **2011**, *115*, 2965–2974.
20. Holopainen, M.; Vastaranta, M.; Kankare, V.; Rätty, M.; Vaaja, M.; Liang, X.; Yu, X.; Hyyppä, J.; Hyyppä, H.; Viitala, R.; others. Biomass estimation of individual trees using stem and crown diameter TLS measurements. *ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* **2011**, *3812*, 91–95.

21. Kankare, V.; Joensuu, M.; Vauhkonen, J.; Holopainen, M.; Tanhuanpää, T.; Vastaranta, M.; Hyyppä, J.; Hyyppä, H.; Alho, P.; Rikala, J.; others. Estimation of the timber quality of scots pine with terrestrial laser scanning. *Forests* **2014**, *5*, 1879–1895.
22. Pokharel, B.; Groot, A.; Pitt, D.G.; Woods, M.; Dech, J.P. Predictive Modeling of Black Spruce (*Picea mariana* (Mill.) BSP) Wood Density Using Stand Structure Variables Derived from Airborne LiDAR Data in Boreal Forests of Ontario. *Forests* **2016**, *7*, 311.
23. Blanchette, D.; Fournier, R.A.; Luther, J.E.; Côté, J.F. Predicting wood fiber attributes using local-scale metrics from terrestrial LiDAR data: A case study of Newfoundland conifer species. *Forest Ecology and Management* **2015**, *347*, 116–129.
24. Hilker, T.; Frazer, G.W.; Coops, N.C.; Wulder, M.A.; Newnham, G.J.; Stewart, J.D.; van Leeuwen, M.; Culvenor, D.S. Prediction of wood fiber attributes from LiDAR-derived forest canopy indicators. *Forest Science* **2013**, *59*, 231–242.
25. Luther, J.E.; Skinner, R.; Fournier, R.A.; van Lier, O.R.; Bowers, W.W.; Côté, J.F.; Hopkinson, C.; Moulton, T. Predicting wood quantity and quality attributes of balsam fir and black spruce using airborne laser scanner data. *Forestry* **2013**, *87*, 313–326.
26. Maltamo, M.; Peuhkurinen, J.; Malinen, J.; Vauhkonen, J.; Packalén, P.; Tokola, T.; others. Predicting tree attributes and quality characteristics of Scots pine using airborne laser scanning data. *Silva Fennica* **2009**, *43*, 507–521.
27. Van Leeuwen, M.; Hilker, T.; Coops, N.C.; Frazer, G.; Wulder, M.A.; Newnham, G.J.; Culvenor, D.S. Assessment of standing wood and fiber quality using ground and airborne laser scanning: a review. *Forest Ecology and Management* **2011**, *261*, 1467–1478.
28. Schütt, C.; Aschoff, T.; Winterhalder, D.; Thies, M.; Kretschmer, U.; Spiecker, H. Approaches for recognition of wood quality of standing trees based on terrestrial laserscanner data. *Laser-scanners for forest and landscape assessment. Proceedings of the ISPRS working group VIII/2, Freiburg, Germany. International Archives of Photogrammetry, Remote Sensing, and Spatial Information Sciences* **2004**, *36*, 179–182.
29. Thomas, L.; Shaffer, C.A.; Mili, L.; Thomas, E. Automated detection of severe surface defects on barked hardwood logs. *Forest products journal* **2007**, *57*, 50.
30. Kretschmer, U.; Kirchner, N.; Morhart, C.; Spiecker, H. A new approach to assessing tree stem quality characteristics using terrestrial laser scans. *Silva Fennica* **2013**, *47*, 1–14.
31. Thomas, L.; Thomas, E.; others. A graphical automated detection system to locate hardwood log surface defects using high-resolution three-dimensional laser scan data. Proceedings of the 17th Central Hardwood Forest Conference, 2010, pp. 5–7.
32. Nguyen, V.T.; Kerautret, B.; Debled-Rennesson, I.; Colin, F.; Piboule, A.; Constant, T. Segmentation of defects on log surface from terrestrial lidar data. Pattern Recognition (ICPR), 2016 23rd International Conference on. IEEE, 2016, pp. 3168–3173.
33. Colin, F.; Mecherqui, R.; Dhôte, J.F.; Fontaine, F. Epicormic ontogeny on *Quercus petraea* trunks and thinning effects quantified with the epicormic composition. *Annals of forest science* **2010**, *67*, 813–813.
34. Bishop, C.M. *Neural networks for pattern recognition*; Oxford university press, 1995.
35. Cortes, C.; Vapnik, V. Support-vector networks. *Machine learning* **1995**, *20*, 273–297.
36. Breiman, L. Random forests. *Machine learning* **2001**, *45*, 5–32.
37. Devroye, L.; Györfi, L.; Lugosi, G. *A probabilistic theory of pattern recognition*; Vol. 31, Springer Science & Business Media, 2013.
38. Terzopoulos, D.; Fleischer, K. Deformable models. *The visual computer* **1988**, *4*, 306–331.
39. Niemeyer, J.; Rottensteiner, F.; Soergel, U. Contextual classification of lidar data and building object detection in urban areas. *ISPRS journal of photogrammetry and remote sensing* **2014**, *87*, 152–165.
40. Díaz-Uriarte, R.; De Andres, S.A. Gene selection and classification of microarray data using random forest. *BMC bioinformatics* **2006**, *7*, 3.
41. Chehata, N.; Guo, L.; Mallet, C. Airborne lidar feature selection for urban classification using random forests. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* **2009**, *38*, W8.
42. Guo, L.; Chehata, N.; Mallet, C.; Boukir, S. Relevance of airborne lidar and multispectral image data for urban scene classification using Random Forests. *ISPRS Journal of Photogrammetry and Remote Sensing* **2011**, *66*, 56–66.

43. Yuan, F.; Sawaya, K.E.; Loeffelholz, B.C.; Bauer, M.E. Land cover classification and change analysis of the Twin Cities (Minnesota) Metropolitan Area by multitemporal Landsat remote sensing. *Remote sensing of Environment* **2005**, *98*, 317–328.
44. Rodriguez-Galiano, V.F.; Ghimire, B.; Rogan, J.; Chica-Olmo, M.; Rigol-Sanchez, J.P. An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing* **2012**, *67*, 93–104.
45. Gislason, P.O.; Benediktsson, J.A.; Sveinsson, J.R. Random forests for land cover classification. *Pattern Recognition Letters* **2006**, *27*, 294–300.
46. Mutanga, O.; Adam, E.; Cho, M.A. High density biomass estimation for wetland vegetation using WorldView-2 imagery and random forest regression algorithm. *International Journal of Applied Earth Observation and Geoinformation* **2012**, *18*, 399–406.
47. Othmani, A.; Voon, L.F.L.Y.; Stolz, C.; Piboule, A. Single tree species classification from Terrestrial Laser Scanning data for forest inventory. *Pattern Recognition Letters* **2013**, *34*, 2144–2150.
48. Rosin, P.L. Unimodal thresholding. *Pattern recognition* **2001**, *34*, 2083–2096.
49. Breiman, L.; Friedman, J.; Stone, C.J.; Olshen, R.A. *Classification and regression trees*; CRC press, 1984.
50. Bradski, G. The OpenCV Library. *Dr. Dobb's Journal of Software Tools* **2000**.
51. Hu, M.K. Visual pattern recognition by moment invariants. *IRE transactions on information theory* **1962**, *8*, 179–187.
52. Wold, S.; Esbensen, K.; Geladi, P. Principal component analysis. *Chemometrics and intelligent laboratory systems* **1987**, *2*, 37–52.
53. DGtal: Digital Geometry tools and algorithms library. <http://dgtal.org>.
54. Tague-Sutcliffe, J. The pragmatics of information retrieval experimentation, revisited. *Information Processing & Management* **1992**, *28*, 467 – 490. Special Issue: Evaluation Issues in Information Retrieval.
55. Provost, F.; Kohavi, R. Glossary of terms. *Journal of Machine Learning* **1998**, *30*, 271–274.
56. Nguyen, V.T.; Kerautret, B.; Debled-Rennesson, I.; Colin, F.; Piboule, A.; Constant, T. Algorithms and Implementation for Segmenting Tree Log Surface Defects. International Workshop on Reproducible Research in Pattern Recognition. Springer, 2016, pp. 150–166.
57. Puissant, A.; Rougier, S.; Stumpf, A. Object-oriented mapping of urban trees using Random Forest classifiers. *International Journal of Applied Earth Observation and Geoinformation* **2014**, *26*, 235–245.

© 2018 by the authors. Submitted to *Remote Sens.* for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).



## Chapitre 4

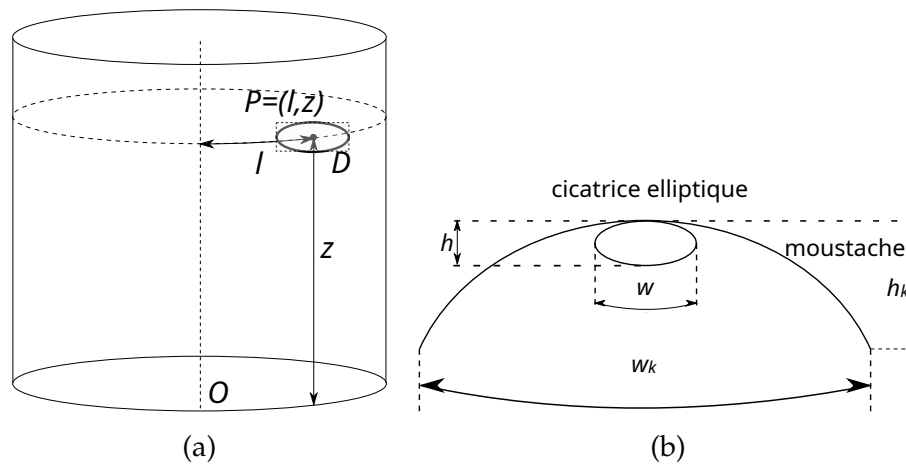
# Caractérisation des défauts externes : vers une évaluation automatique de la qualité de l'arbre sur pied

### 4.1 Contexte et objectifs

Le classement des arbres sur pied est une tâche essentielle de l'inventaire forestier et pour préparer la récolte (Fonseca, 2005). Une procédure de classement consiste à mesurer des attributs globaux du tronc tels que le volume et la courbure et à quantifier les défauts. Les défauts locaux sont parmi les facteurs les plus critiques qui influencent la qualité du bois et la transformation ultérieure du bois. Ces défauts peuvent dégrader les qualités aussi bien mécaniques qu'esthétiques limitant l'utilisation du bois. Dans tous les cas, les défauts déprécient la valeur marchande.

Dans les deux chapitres précédents, nous avons développé avec succès la détection des défauts de surface avec des dimensions pouvant descendre jusqu'à 0,5 cm et ensuite nous avons exploré l'identification en utilisant une approche d'apprentissage automatique, basée sur les Random Forests (Breiman, 2001). Les résultats de la classification des défauts sur les troncs de plusieurs espèces d'arbres telles que le chêne, le hêtre, le merisier, le pin, le sapin et l'épicéa sont apparus prometteurs. La valeur moyenne de la mesure  $F_1$  de la classification était de 0,84. Les résultats de la détection et de la classification ont montré une bonne robustesse par rapport aux caractéristiques des espèces et des arbres telles que la rugosité de l'écorce, l'âge, le diamètre et la forme du tronc et ceci pour des défauts différents. Ces résultats nous permettent d'aller plus loin pour qualifier les défauts détectés et pour classer l'arbre sur pied.

L'objectif de ce chapitre est de tirer profit de la quantification des défauts détectés et classés par les méthodes décrites dans les deux chapitres précédents pour développer une application



**Figure 4.1** Description du calcul de la position du défaut  $D$  par les coordonnées de son centre  $P$ , les dimensions de  $D$  sont déterminées par sa boîte englobante (a). Une cicatrice de branche se caractérise par une cicatrice elliptique et une moustache (b).

permettant d'évaluer automatiquement la qualité de l'arbre sur pied. Le résultat de ce chapitre a fait l'objet d'une publication (section 4.6).

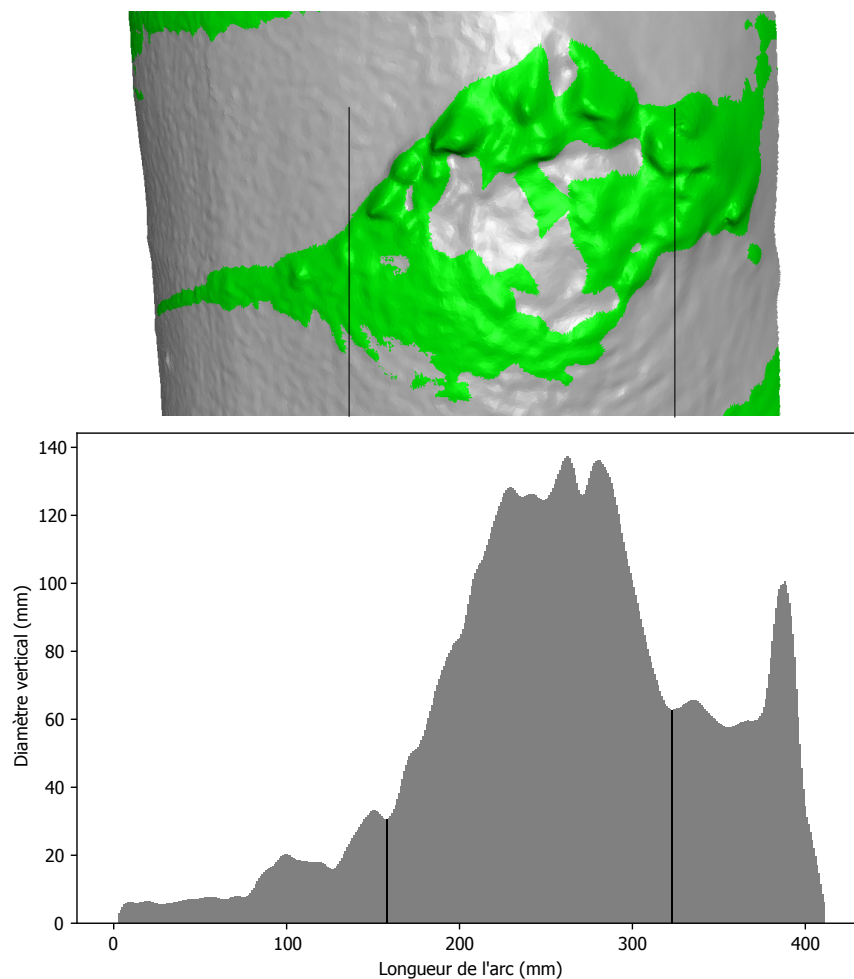
## 4.2 Méthodologie

Le nuage de points issu des scans TLS du tronc est d'abord traité pour réduire le bruit et pour reconstruire la surface du tronc en créant un maillage. Le processus principal se décompose ensuite en 3 étapes dont la détection et l'identification des défauts, présentées dans les deux chapitres précédents. L'objectif principal est maintenant de déduire les attributs des défauts détectés qui sont leur position et leurs dimensions.

La position d'un défaut sur le tronc peut être déduite directement à partir des coordonnées des points du défaut. Afin de nous approcher des méthodes de mesure utilisées par un opérateur sur le terrain, une ligne de référence est définie sur l'écorce en suivant la direction globale du tronc. Le défaut est représenté dans un système de coordonnées  $l, z$  où  $l$  est la longueur de l'arc (perpendiculaire à la ligne de référence) entre le centre du défaut et la ligne de repère, et  $z$  est la distance entre le défaut et le sol (Figure 4.1 (a)). La position du défaut est calculée par le centre de la boîte englobante de la jonction entre le défaut et le tronc (Figure 4.1 (a)). La boîte englobante d'un objet est la boîte la plus petite qui contient l'objet.

Les dimensions d'un défaut sont estimées différemment selon son type qui à ce stade est connu.

- Pour un défaut de type branche, les paramètres sont son diamètre et son angle d'insertion. Nous avons utilisé l'approche de la ligne centrale pour estimer le diamètre de branche. Le diamètre est calculé par le mode (la valeur la plus fréquente) de la distance à la ligne centrale de la branche. L'angle d'insertion peut aussi être déduire à partir de la ligne centrale de



**Figure 4.2** Illustration du calcul des dimensions de la cicatrice de la branche en utilisant le profil des extensions verticales du défaut le long de l'arc périphérique.

la branche. Cependant, nous n'avons pas de données de référence pour comparer avec la mesure automatique.

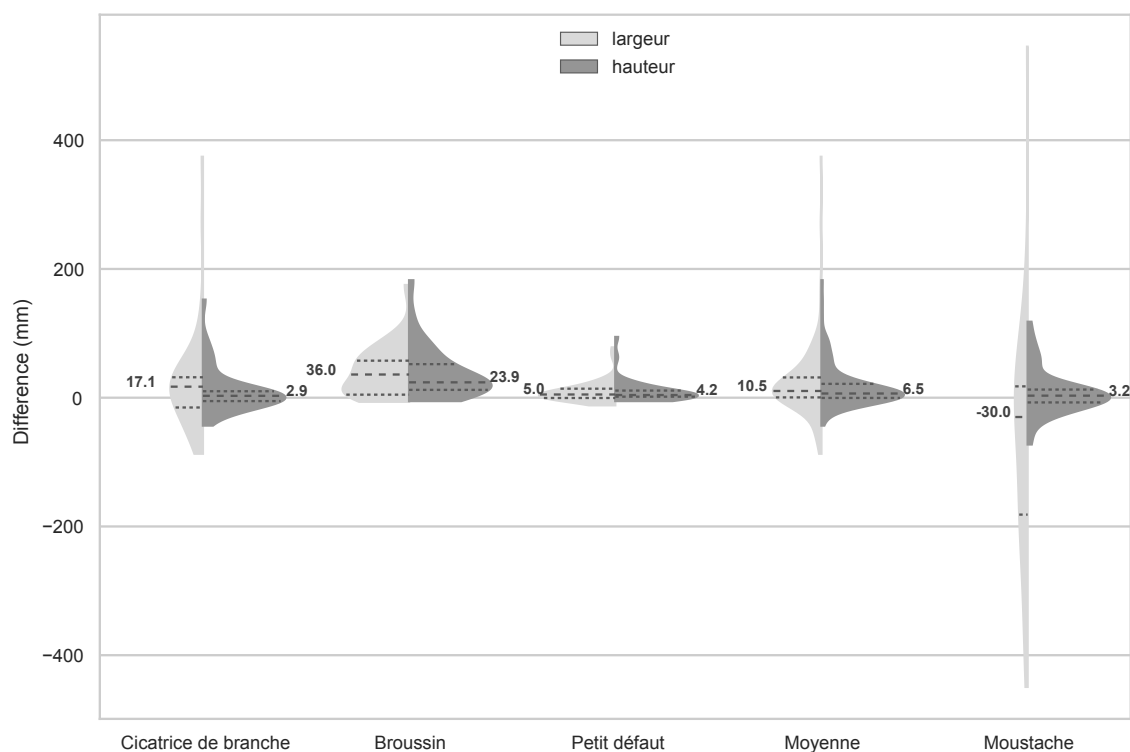
- Une cicatrice de branche (sur une écorce lisse) se caractérise par une cicatrice elliptique et une moustache. L'estimation des dimensions d'un défaut de ce type consiste donc à estimer les diamètres (horizontal et vertical) de la cicatrice elliptique et les dimensions de la moustache incluant la largeur et la hauteur (Figure 4.1 (b)). Les dimensions de la moustache sont déduites directement de la boîte englobante du défaut. Les dimensions de la cicatrice sont plus difficiles à déduire. Nous avons proposé une méthode basée sur le profil périphérique de l'extension verticale de la cicatrice de branche et qui fait l'hypothèse que le maximum de l'extension correspond au petit diamètre de l'ellipse (Figure 4.2). Pour plus de détails, voir la section 2.5.2 de l'article 5 (section 4.6).
- Les défauts d'autres types tels que les broussins, les amas de bourgeons, les picots et les sphéroblastes sont caractérisés par leurs diamètres horizontal et vertical. Ces diamètres sont estimés par la boîte englobante de la jonction entre le défaut et la surface du tronc.

A partir des informations sur la position, le type et les dimensions des défauts détectés, nous pouvons évaluer la qualité des troncs sur pied selon une norme définie en fonction de l'espèce de l'arbre (AFNOR, 2012a; Jourez et al., 2010).

Les résultats de l'estimation automatique des dimensions des défauts ont été comparés avec les mesures de référence qui avaient été faites avant que les scans aient été réalisés.

### 4.3 Résultats principaux

L'expérimentation sur la mesure des dimensions et position des défauts a été menée sur 3 chênes et 3 hêtres. Le Random Forests entraîné dans l'étape d'identification a été utilisé pour identifier les défauts avant la mesure. Les résultats de l'expérimentation sont apparus prometteurs. Après les étapes de détection et d'identification, notre programme a reconnu correctement 121 sur 141 défauts avec 85 faux positifs dont 82% sont liés à de très faibles irrégularités de l'écorce sur le hêtre telles que le *Nectria* et la plissure d'écorce. Nous avons comparé les dimensions estimées de ces 121 défauts avec les dimensions observées. La Figure 4.3 montre les diagrammes en violon des différences entre l'estimation et l'observation des dimensions du défaut pour différents types de défauts.



**Figure 4.3** Diagrammes en violon des différences entre l'estimation et l'observation des dimensions du défaut pour différents types de défauts. Pour chaque diagrammes, les lignes horizontales indiquent les quartiles et la médiane, cette dernière étant accompagnée par sa valeur.



En général, l'algorithme a fourni un bon résultat avec une médiane proche de zéro tant pour la largeur (10,5 mm) que pour la hauteur (6,5 mm). Nous constatons que les dimensions des défauts de tous les types ont été surestimées. A partir de ces diagrammes, nous pouvons voir que l'estimation des hauteurs (diamètres verticaux) des cicatrices elliptiques était plus fiable que celle de leurs largeurs (diamètres horizontaux), l'estimation des largeurs présente même quelques données aberrantes. Les diagrammes des moustaches des cicatrices de branche montre une grande dispersion. Leur largeur est difficile à estimer et est souvent sous-estimées (médiane égale -30 mm) car les moustaches étaient souvent fines notamment aux bouts de leurs branches qui sont elles-mêmes difficiles à entièrement détecter. Sur les broussins, les médianes de la différence entre les dimensions estimées et observées sont plus grandes mais la dispersion reste assez faible. Les petits défauts ont une distribution des différences plus compactes et avec des médianes proches de zéro (5,0 mm pour la largeur et 4,2 mm pour la hauteur).

La performance de notre algorithme d'estimation automatique de la hauteur des cicatrices de branche est comparable à la mesure semi-automatique de Kretschmer et al. (2013). Le pourcentage des défauts avec une différence entre les diamètres verticaux estimés et observés de moins de 1 cm est de 56 %, très proche de la valeur de 58 % obtenue par Kretschmer et al. (2013).

N'étant pas la cible principale de cette étude, focalisée sur les défauts de faible relief, seules deux branches figurent dans l'échantillon de défauts. Néanmoins, la méthode de calcul mise en œuvre fournit dans ces deux cas des résultats corrects en dépit d'une courbure forte des insertions et d'une quantité importante de bruit caractéristique de ces jonctions, et semble bien adaptée à la caractérisation voulue d'un départ de branche.

## 4.4 Intégration à la plate-forme Computree

L'objectif plus appliqué et de valorisation final du projet était de fournir un outil informatique intégrant nos développements en vue de classer des troncs ou des grumes qui s'intégrait dans l'environnement de la plate-forme Computree (<http://computree.onf.fr>), un logiciel "open source" dédié au traitement des données LiDAR en foresterie. Nos développements sont disponibles sous la forme d'un plugin de Computree et testés dans un système d'exploitation Linux (Figure 4.4). A ce stade, il reste à assurer la compatibilité avec d'autres systèmes d'exploitation tels que Windows notamment en ce qui concerne l'intégration de bibliothèques externes telle que DGtal (<https://dgtal.org>).

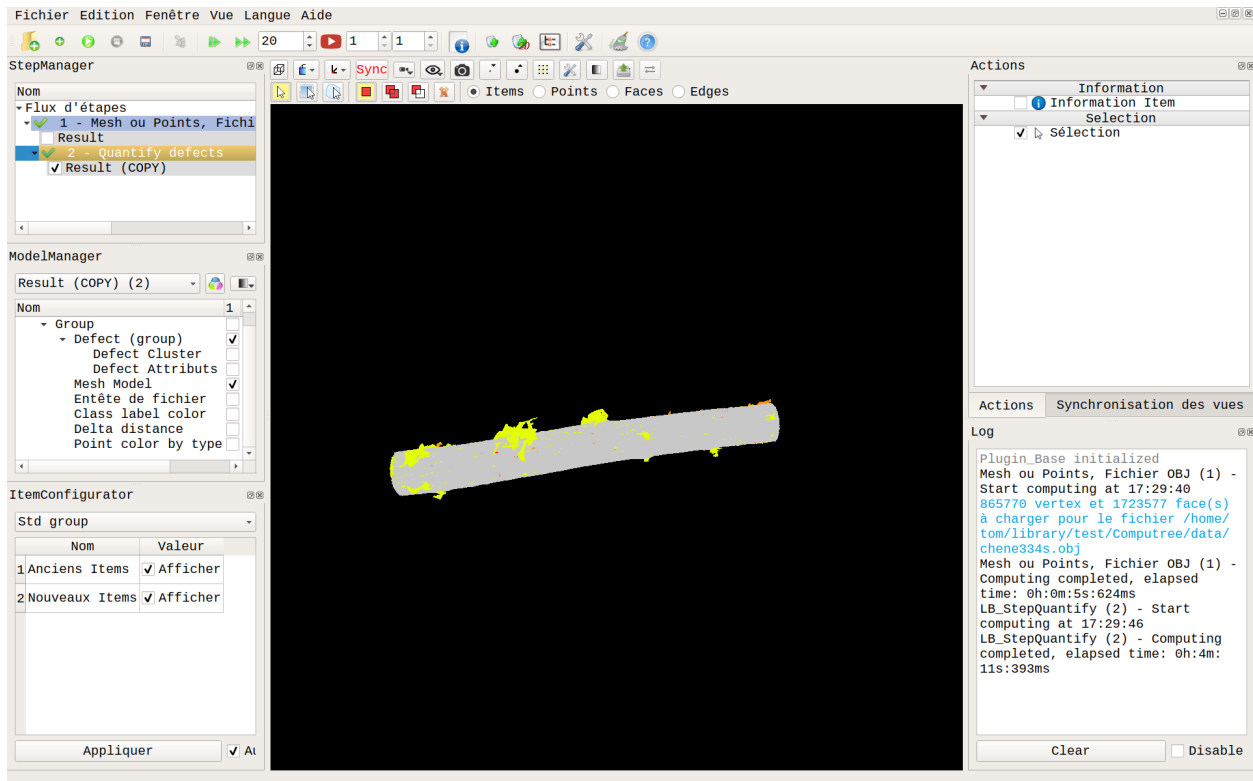


Figure 4.4 Capture d'écran du plugin dans le plateforme Computree

## 4.5 Conclusion

Ce chapitre a présenté une méthode pour caractériser les défauts déjà détectés et identifiés durant les étapes précédentes. Ce chapitre complète la suite des algorithmes pour finalement tester l'évaluation de la qualité du bois à partir de données TLS, en confrontant les caractéristiques des défauts obtenues à partir de l'information 3D à des mesures manuelles de ces caractéristiques.

Ce travail propose une approche automatisée pour quantifier les défauts de tous les types liés à la ramification. En comparant avec les mesures manuelles, les mesures automatisées fournissent souvent un nombre de défauts plus nombreux à cause de faux positifs correspondant souvent à de faibles variations du relief sur des écorces lisses qui ne sont pas considérées comme des défauts par l'opérateur humain du fait de son expertise. L'autre différence majeure provient de la délimitation du défaut qui diffère selon l'opérateur humain ou automatique et conduit à des dimensions souvent surestimées en automatique. La distribution des différences entre les dimensions estimées et observées est assez dispersée. Toutefois, pour une première tentative de comparaison, les résultats sont prometteurs et justifient de poursuivre et de continuer à améliorer les différentes étapes permettant d'arriver à ce résultat en bout de chaîne de traitement. Comme souvent avec cette technologie, la validation est délicate car elle pose la question de disposer d'une mesure de référence qui ne se base pas sur la même source d'information. Néanmoins de nombreuses améliorations

possibles ont déjà été discutées dans les chapitres précédents et seront résumées dans le chapitre 5.

## **4.6 Article 5 : An innovative and automated method for characterizing wood defects on trunk surface using high density 3D terrestrial LiDAR data**

Article soumis pour publication dans la revue *Annals of forest science*

## **An innovative and automated method for characterizing wood defects on trunk surface using high density 3D terrestrial LiDAR data**

**Key message** We designed a novel method allowing to automatically detect defects on the surface of tree trunk including branches, branch scars, and epicormics from terrestrial Lidar data by using only 3D information. We could detect the defects with a diameter as small as 0.5 cm on either oak or beech trees, the former having a rough bark and the latter having a smooth bark.

### **Abstract**

**Context:** The ground-based Light Detection and Ranging (LiDAR) technology provides a description of standing trees with a high level of details. This gives the opportunity to assess standing tree quality and to use this information in forest inventory. Several tree attributes affect standing tree quality. Among them, branching defects deserve a specific attention.

**Aims:** This study aims to propose a computing method to automatically detect, identify and quantify the defects on the trunk surface using the 3D data recovered from terrestrial LiDAR scanning. A second objective is to validate the efficiency of this technology by comparing its results with manual measurements.

**Methods:** We proposed a three-step approach. The potential defects on the trunk surface were first detected by a segmentation algorithm. Second, the defect type was identified using a random forest machine learning algorithm allowing the elimination of false detections. Third, for each detected defect we estimated its position, horizontal, and vertical dimensions. Finally, the characteristics of the defects were compared to those observed directly on the trunk by an operator.

**Results:** The defects were detected and classified with a good rate with an average  $F_1$  score of 0.8. Despite differences between computed and observed defect areas resulting from the detection methods, the results are very encouraging, as shown by comparisons between the number, position, and size of defects. The implementation of the suite of algorithms is a real improvement on previous methods.

### **Conclusion**

The information about the defects present on the trunk surface measured from TLS data can be used in an automated procedure for grading standing trees or roundwoods.

**Keywords:** tree quality, roundwood quality, forest inventory, wood grading, 3D data

## 1. Introduction

Grading standing trees is an essential task in forest inventory and pre-harvest (Fonseca 2005). Beside the global attributes such as volume and curvature, the local defects are among the most critical factors influencing the wood quality and the subsequent wood processing. These defects can degrade the mechanical performance, leading to high security coefficients for structural use or the aesthetics for interior or furniture uses. In any case, they depreciate the economic value. The best research method for evaluating standing tree quality is to climb on the tree using a ladder and note all the defects and their attributes such as position, dimension, and type (Colin et al. 2010b), but for inventory purposes a visual inspection from the ground is generally completed (Jourez, et al. 2010). For grading roundwood, X-ray tomography can be used to measure the knots and other defects inside a log. Although this approach provides the best performance, its main drawbacks are the requirement for felling the tree and very high investment costs (Thomas et al. 2006).

Lidar (Light Detection and Ranging) can measure an object in 3D dimensions using a technique whereby a laser beam is emitted and the reflected light received by the detector. The data provided by Lidar is a point cloud, which contains the three spatial coordinates of the intercepted points and additional characteristics such as reflectance or RGB values, which are not used in this work. A wide range of forestry applications has been developed recently (Popescu et al. 2002; Simonse et al. 2003; Hopkinson et al. 2004; Thies and Spiecker 2004; Henning and Radtke 2006; Litkey et al. 2008; Holopainen et al. 2011; Yao et al. 2011; Dassot et al. 2012; Mutanga et al. 2012; Kankare et al. 2014). In particular, many studies have taken advantage of the potential of Lidar for replacing conventional methods for measuring forest inventory parameters such as tree height, diameter at breast height (DBH, trunk diameter measured at 1.3 m above ground level), density per hectare of the number of stems, and volume of wood for assessing biomass.

The Terrestrial LiDAR technology (TLS) has been increasingly used for characterizing not only forest stands but also individual standing trees thanks to the high 3D details (Holopainen et al. 2011; Dassot et al. 2012; Kankare et al. 2013, 2014; Calders et al. 2015; Hackenberg et al. 2015; Wang et al. 2016; You et al. 2016). In particular, the high-density data provided by TLS allows an accurate reconstruction of the tree bark surface. The analysis of the bark pattern can be then used to recognize tree species (Othmani et al. 2013) or to detect the local defects on the bark (Schütt et al. 2004; Thomas et al. 2006, 2010; Kretschmer et al. 2013; Stängle et al. 2013; Nguyen et al. 2016). The quality of standing trees or roundwoods can be inferred by using this detection. The TLS is an alternative to the manual measurement of standing trees or to an X-Ray tomography measurement of knots inside logs of felled trees. In fact, there is a strong correlation between the external information and the internal information about wood knots (Thomas 2009; Stängle et al. 2013; Racko 2013; Pyörälä et al. 2017) with a coefficient of determination  $R^2$  generally greater than 0.6. There are, however, few studies about grading standing trees or roundwoods using TLS data (Schütt et al. 2004; Thomas et al. 2006, 2010; Kretschmer et al. 2013; Stängle et al. 2013). In order to detect defects, most approaches used the difference between the surface of the trunk and a reference surface resulting from a primitive shape fitting such as circle (Thomas et al. 2006, 2010) or cylinder (Schütt et al. 2004; Kretschmer et al. 2013; Stängle et al. 2013). Because the cross-sections of a log are often not perfectly round, the primitive shapes hardly fit the real log surface. The methods based on this technique thus detected only the large defects with a diameter higher than 12.7 cm and protruding more than 2.5 cm from log surface (Thomas et al. 2006). Depending on species and tree age, the bark may have different structures (smooth or furrowed) that can interfere with the detection of defects. Therefore, an automated process is challenging to

design, especially when one wants to detect the small defects. The previous methods on the estimation of tree quality using TLS were mostly semi-automatic (Schütt et al. 2004; Kretschmer et al. 2013; Stängle et al. 2013) and limited to large branch scar defects. Schütt et al. (2004) presented a semi-automatic approach to detect and classify wood defects using both range and intensity information of TLS data. The potential defects were firstly segmented by the region based segmentation of grey scale image derived from the distance between 3D point cloud of the log and the fitted cylinders. A manual adjustment can be done to make the final decision about these potential defects that can be then classified by a neural network. However, neither information about how to train the neural network nor the results were presented. Thomas et al. (2006, 2010) proposed a method to automatically detect severe defects on red oak and yellow poplar using a multi-scale analysis of contours from 2D images derived from the distance between 3D points describing the log and the fitted circles. More recently, Kretschmer et al. (2013) used a 3D approach by using a distance map of the difference between the trunk surface and the reference surface obtained by fitting cylinders. The highlight in the distance map was used as a visual aid for the user to detect and measure branch scar on European beech within a diameter range from 4 cm to 18.2 cm. A similar method was used by Stängle et al. (2013) to estimate the knotty core size in European beech. The comparison with the X-Ray scanning results led to an agreement of 62.5% between both grading methods.

While branch scar is indisputably one of the most common defects on trunks with a strong impact on the wood quality, other defects, such as burl, picot (Colin et al. 2010a) must not be neglected as mentioned in the standards (Carpenter and Jones 1989; AFNOR 1997; Jourez et al. 2010). For example, epicormics consisting in burl, picot and bud cluster, may degrade the quality of oak at least one class (Meadows and Burkhardt 2001).

In previous works, we successfully developed the detection of the surface defects with dimensions as small as 0.5 cm (Nguyen et al. 2016) and then classify them by using a machine learning approach (Nguyen et al. 2018). The feature vector used by the random forests included the species, Hu moment invariants (Hu 1962), dimension, eigenvalues, and mean of roughness of the defect point cloud. The results obtained on trunks of several tree species including oak, beech, wild cherry, pine, fir, and spruce were rather promising with an average  $F_1$  score of 0.835. The results showed a robustness to the tree species features, such as furrowed bark, age, trunk diameter and shape. These promising results allow us to go to a further step to estimate defects attributes.

In the context of tree quality assessment, our first objective was to use our previously designed algorithms including the detection and classification of defects for producing a complete suite of algorithms up to the automatic delivering of their attributes including position and dimension. The second objective was to verify the efficiency of the whole data processing workflow for grading standing trees by comparing the estimated defect attributes with their manual measurements.

## 2. Material and methods

### 2.1. Method overview

We developed a method consisting in three steps (Figure 1). The first step “Detection” aimed to quickly find the smallest as possible candidate defects on the mesh of the trunk. A threshold-based segmentation algorithm was developed to detect the occurrence of a defect based on the difference between the reference distance and the distance of 3D points to the trunk centerline. The second step “Identification” had two main objectives: 1) to remove the false positives coming from the first step and 2) to classify the candidate defects into different types

by their biological nature. The last step consisted of estimating the position and dimensions of the identified defects.

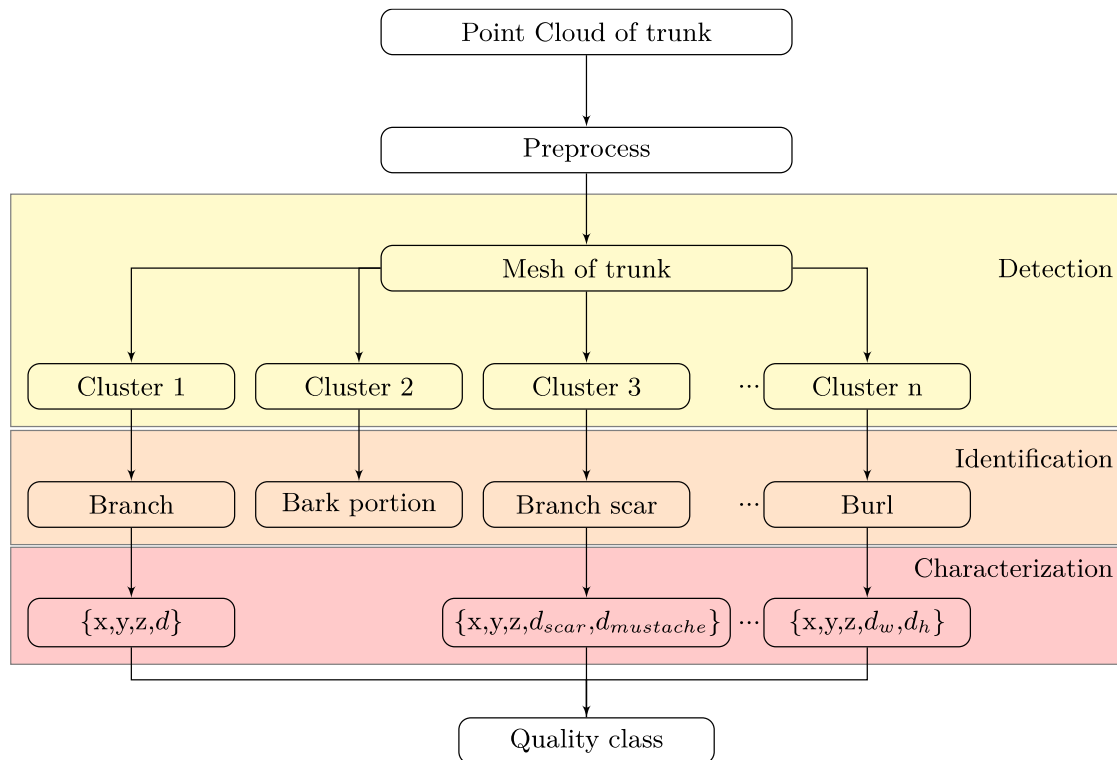


Figure 1: Overview of the method divided into three steps.

## 2.2. Data collecting and preprocessing

The samples were measured with a scanner Faro Focus 3D X130 in the Champenoux forest close to Nancy, France (48°43'30.1"N 6°20'52.1"E). The database consists in point clouds of trees of two species, oak (*Quercus petraea*) and European beech (*Fagus sylvatica*). In order to detect the small defects, we used a high-resolution setting and put the scanner close to the trunk. The resolution was the half of the maximum value, and the distance from the scanner to the trunk was from 3 m to 4 m. With this setting, the horizontal and vertical angular step size of the scans were 0.0045°, which results in a point spacing of 0.31 mm at a distance of 4 m. Each tree was scanned from 2 points of view targeting towards the same side. The two scans were then merged into a unique file to recover the 3D view of one side of the trunk by using the Faro Scene software.

The defects with a diameter greater or equal to 0.5 cm were described prior to carrying out the scans in terms of dimension, type, and position. The position was taken in a common coordinate system  $(h,l)$  where  $h$  is the height and  $l$  is the arc length from the reference line to the defect. The reference line was a measuring tape placed at the center of the scanned side of the trunk. It was also used to measure the  $h$  coordinate. The ruler was removed before scanning the tree. Two small markers (ping-pong balls or small pieces of paper) at 1 m height and 5 m height were kept in the scans for a synchronization with the automatic measurement. The area of the evaluation was from 1 m to 5 m in trunk length and a half of tree circumference.

The manual measurement was performed before carrying out the scans. Firstly, the position of the defect was measured in the  $(l,z)$  coordinate system (Figure 2 (a)). The origin of the coordinate system  $O$  was located at 1m



above the soil and at the center of the scanning tree side which was marked by a ping-pong ball or a piece of paper. The  $z$  coordinate was the central axis of the trunk. The  $l$  coordinate was the arc length from the central axis. The measurement was carried out on one side of the tree with  $l$  ranging from  $-\frac{1}{4}$  circumference to  $\frac{1}{4}$  circumference at Breast height and with the height  $z$  ranging from 1 m to 5 m from the soil. The measured defects were limited to a diameter of 5 mm without considering the limit between the protuberance and the surrounding bark.

The measurement of the defect dimension depends on the defect type. For burl and small defects, the measurements were the defect width  $w$  and height  $h$  (Figure 2 (b)). The width  $w$  was measured by the arc length between the leftmost and rightmost points of the intersection between the defect and the bark (see Figure 2 (b)). The height  $h$  was measured by the difference between the maximum and minimum  $z$  coordinates of the intersection between the defect and the bark. For branch scar, in addition to the width  $w$  and height  $h$  of the scar, the mustache width ( $w_k$ ) and height ( $h_k$ ) were also measured by the same method used to measure  $w$  and  $h$  respectively (Figure 2 (c)). For the defects of the branch type including both sequential and epicormic branches, the branch diameter was measured at a short distance from the trunk after the junction between the branch and the trunk (Figure 2 (d)).

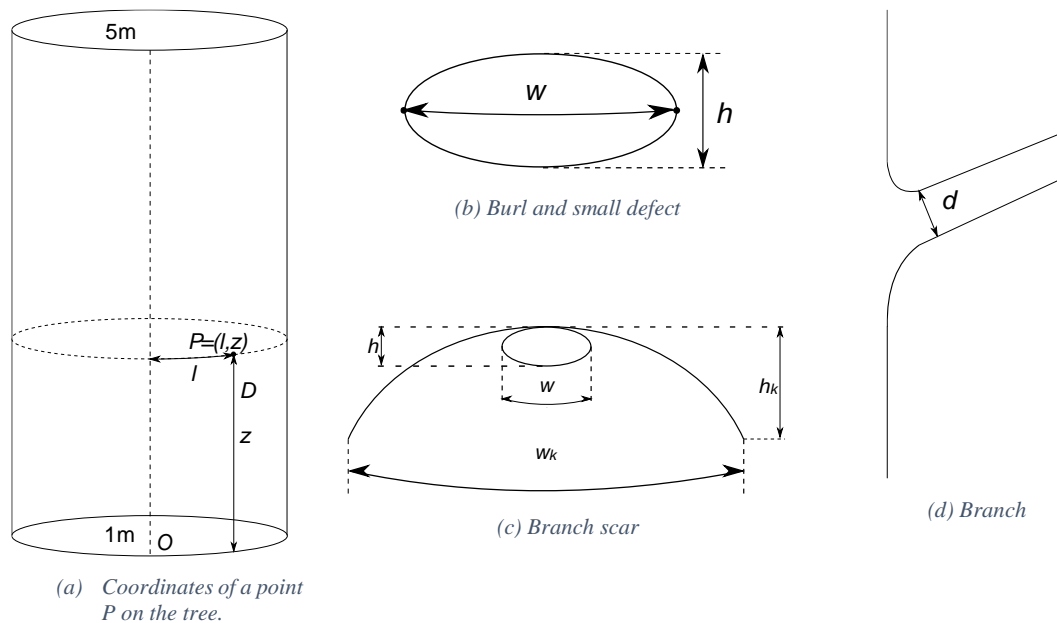


Figure 2: Manual measurement of defect width and height for different defect types.

Six trees including three oak trees and three beech trees were used to validate the method. In addition to these data, other data were used to train the Random Forest (Table 1).

Table 1: Number and attributes of the sample trees

Species	Nb of trees used to validate the measurement	Nb of trees used to train the Random Forest	Diameter at Breast Height(cm)
Oak	3	6	35-76
Beech	3	5	30-57
Wild Cherry		2	20-26
Pine		1	34
Fir		2	23-42
Total	6	16	

The point cloud was preprocessed to enhance the outcome of the process. The noise was reduced by using our algorithm based on Euclidian distance clustering (Nguyen et al. 2018). Then the point cloud with its reduced noise was processed in the Graphite software (<http://alice.loria.fr/index.php/software.html>) including the following steps : (1) smoothing of the point cloud, (2) meshing of the point cloud and (3) remeshing the result of step (2) to obtain a smooth mesh. More details regarding the process were presented in Nguyen et al. (2018).

### 2.3. Detection

The detection of the defects is based on an enhanced version (Nguyen et al. 2018) of an initial segmentation algorithm (Nguyen et al. 2016). The outer part of a defect is typically a protruded region comparing to the surface of the normal area around it. First, we computed the centerline of the trunk using an adaptation of the method proposed in Kerautret et al. (2016). Second, the branches were segmented by dividing the trunk volume into equal cylindrical sectors (pie slices) with height equal  $L$ , and angle equals to  $\frac{L}{R}$  where  $R$  is the radius of the trunk, and  $L$  is an arc length. In each cylindrical sector, the closest point to the centerline will belong to the trunk. We keep all these points as a set of seed points. The points in the point cloud having a distance to at least one seed less than  $\sqrt{2}L$  are inserted in the trunk points set. In most cases, a range of  $L$  between 50 and 100 mm ensures that at least one seed corresponding to the trunk is detected by sector and the  $L$  value can be increased in case of large branches. The algorithm separated the point cloud into two disjointed sets one containing branches and the other the trunk. Third, for each point  $P$  in the set of trunk points, a narrow longitudinal patch consisting of neighbors of  $P$  was defined. The reference distance to the centerline of a point  $P$  was computed from the fitting of the linear trend of all points belonging to the patch by using the least squares method. For each point in the point cloud, we computed the difference between the distance to the centerline and the reference distance to the centerline. An automatic thresholding method (Rosin 2001) was applied on the histogram of this difference for the whole set of trunk points was used to obtain the potential points of defects. The potential defects were obtained by using Euclidean clustering method (Rusu and Cousins 2011) on the union of these potential points of defects and the points of branches.

### 2.4. Identification

As shown in Nguyen et al. (2018), the classification of defects is a critical step since the impact on wood quality depends on the defect type. Furthermore, the relevant characteristics of the defect are type dependent. For example, the branch diameter was measured by the distance to the branch centerline while the burl dimension was measured by the arc length between the leftmost and rightmost points. In Nguyen et al. (2018), we described in details a method based on the supervised machine learning Random Forest (Breiman 2001) using a feature vector included the following features:

1. Species of the tree.
2. Ratio between the number of points of the defect and the volume of its bounding
3. Hu moment Invariants (Hu 1962).
4. Defect width.
5. Ratio between defects width and height.
6. Ratio between defects width and thickness.
7. Mean of the difference between the distance and the reference distance to the centerline of all points of the defect.
8. Standard deviation of the difference between the distance and the reference distance to the centerline of all points of the defect.

9. Ratio between the eigenvalue of the radial vector and the eigenvalue of the longitudinal vector.
10. Ratio between the eigenvalue of the tangential vector and the eigenvalue of the longitudinal vector.
11. Angle between the radial vector and the principal axis of the tree  $\alpha$ .

We used both the manually segmented and the automatically segmented defects to train the Random Forests. The data were labelled prior to the scanning. In addition to the defects types, a bark class was introduced in the Random Forests because it is often miss-classified as a defect by the segmentation algorithm. The classes in our training database included:

1. Branch including sequential branch and epicormics branch
2. Branch scar
3. Burl
4. Small defects including picot, spheroblaste, bud and bud cluster with less than 6 buds.
5. Bark

This Random Forests was used to classify the detected defects. For each candidate defect, its feature vector was computed and then passed to the Random Forests. The Random Forests consists of a set of CARTs (Breiman et al. 1984) (Classification and Regression Trees). Each tree makes a classification, or in other words, gives a vote for a class. The random forest chooses the class which is voted by the majority of trees.

#### 2.5.2 Automatic estimation of the dimensions

We tried to mimic as accurately as possible the approach applied when we manually measured the defects on the standing tree. Consequently, the measurement method depended on the defect type. We needed to represent the defect on a coordinate system similar to what we used on the field. The dimensions of the defects were measured on a cylindrical surface. If the detected defect was a burl, a branch scar or a small defect, it was represented by a cylindrical coordinate system  $(O, r, \theta, z)$ . The origin  $O$  was defined by the intersection of the centerline of the trunk and the plane containing the marker point  $M$  (marked by the small ball on the bark) and this perpendicular to the centerline (Figure 3). The defect was represented in the cylindrical coordinate system by a set of points  $D = \{P(r, \theta, z)\}$  with:

- $r$  the distance from  $P$  to the centerline.
- $z$  the height of the point along the centerline.
- $\theta$  the angle between the radial vector  $\overrightarrow{P'P}$  and  $\overrightarrow{OM}$  with  $P'$  on the centerline, and  $P'P$  being perpendicular to the centerline.

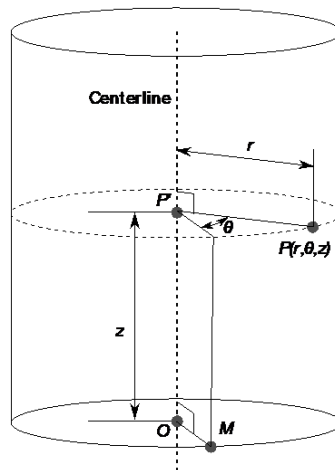


Figure 3: Description of the conversion of a point  $P$  from a Cartesian to a cylindrical coordinate system.

Measurement of the position of the defect and the dimensions  $w$ ,  $h$  of the bur or, small defect and  $w_k$ ,  $h_k$  of branch scar.

In the cylindrical coordinate system, the following procedure allows measuring the width  $w$  and height  $h$  of burls and small defects and the mustache width  $w_k$  and height  $w_k$  of branch scar.

- Compute the minimum distance to the centerline  $r_{min}$
- For a point  $P$  in the defect points
  - o Add  $P$  to the set  $I$  if  $r_p - r_{min} < r_t$ , with  $r_t = 10 \text{ mm}$ ; The threshold  $r_t$  allows to simulating the intersection between the defect and the bark as we do manually on the field.
- Find the bounding box of the set  $I$ , represented by two points  $P_1$  and  $P_2$
- Compute the center of the bounding box of  $I$ :  $P_C = \frac{P_1 + P_2}{2}$
- Compute the radius of the trunk at the defect position  $r_{trunk}$  by using the mode of the distance to the centerline of all the points in a trunk segment of 5 cm length with  $P_C$  is the center of the segment.
- The coordinate  $z$  of the defect is the coordinate  $z$  of  $P_C$ .
- The coordinate  $l$  of the defect is:  $l = \theta_{P_C} * r_{trunk}$
- Compute width of defect  $w$  (or  $w_k$ )
  - o  $\alpha = |\theta_{P_2} - \theta_{P_1}|$
  - o  $w = \alpha * r_{trunk}$
- Compute height of defect  $h$  (or  $h_k$ )
  - o  $h = |z_{P_2} - z_{P_1}|$

Measurement of branch scar dimension  $w$  and  $h$ .

The branch scar can have different shapes such as ellipse or “Chinese” mustache. Since the above-mentioned algorithm was expected to infer the dimensions of the internal knot from the externally visible branch scar, the dimension of the scar that was critical for estimating the knot diameter was more difficult to measure when it has a Chinese mustache shape. We proposed a simple method based on the profile of the height by arc length to detect and measure the branch scar. The idea is that a mustache has a significantly smaller height compared to the scar. Thus, on a profile of height, the boundary between the mustache and the scar can be detected by significant small local minima. The algorithm works with both ellipse and Chinese mustache shape. The algorithm is described as following and illustrated in Figure 4.

- Conversion of the defect points from a cylindrical coordinate system into a  $(l, z)$  coordinate system with
  - o  $l = \theta * r_{trunk}$
- In the  $(l, z)$  coordinate system, compute the  $z$  profile or histogram of the difference between the maximum and minimum  $z$  values of the defect in function of the arc length  $l$ .

- Smoothing the histogram by an average of neighbors using a window of size 7.
- Find the maxima (M) and all local minima.
- Clip the histogram at the first left ( $M_1$ ) and right minima ( $M_2$ ) which have a height less than 40% of the maxima.
- Scar width:  $w = l_{M_2} - l_{M_1}$
- Scar height:  $h = h_M$

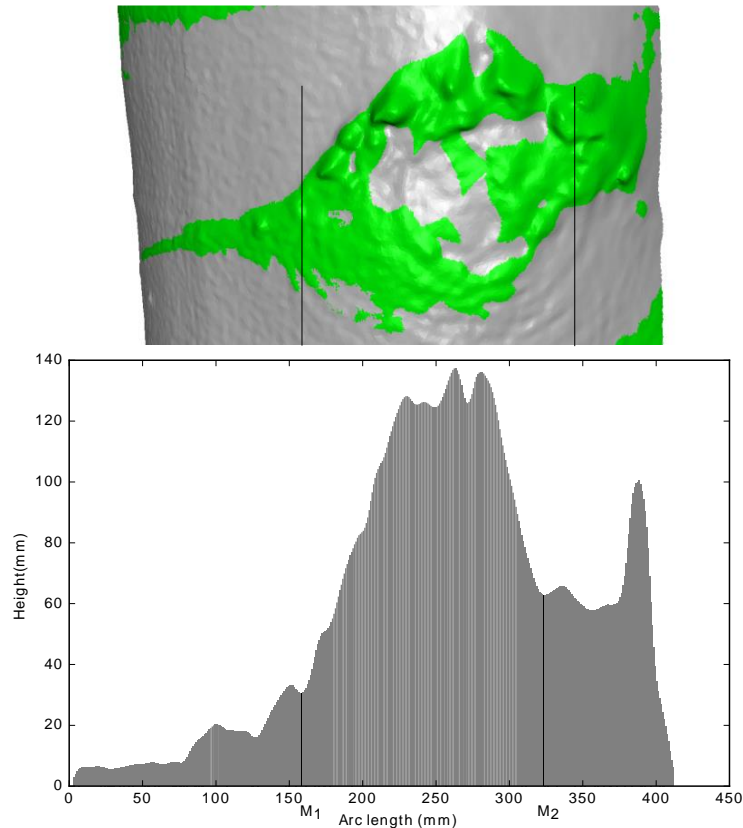


Figure 4: Illustration of the dimensions of the branch scar using height profile by arc length

#### Estimation of branch diameter

The estimation of branch diameter occurred when the classification step returned a branch as defect type. A branch is different from the other defect types because the diameter is not measured at the intersection of the branch and the trunk. We considered that branches have a tubular shape, and two measurement methods were tested. The first method was to fit a cylinder to a portion of the detected branch. The second was to compute the centerline of the branch and compute the mode of the distance to the centerline of points in a portion of the branch close to the trunk. Our experiment showed that the second method was more robust as the branches were often flexuous which impeded the fitting of a cylinder. Moreover, the first method required an *a priori* information of the approximated diameter of the branch for well estimating the true diameter. In fact, the second method was not only able to estimate the branch diameter without *a priori* information of the branch but also robust regarding the noise and the branch flexuosity. The following procedure describes the method based on the centerline.

- 1- A portion of the branch from 3 cm to 8 cm to the trunk was extracted.
- 2- The centerline for the portion was computed by using the same method as for the centerline of the trunk.

- 3- The distance from all points in the portion to the centreline was computed.
- 4- The diameter of the branch was taken as twice the mode of the distance to the centerline.

## 2.6 Implementation of the software

The output of the software is a mesh derived from TLS data of the trunk (the preprocess and creation of the mesh were presented in the section “Preprocessing” or in Nguyen et al. (2018)). The program was implemented in C++ language using several libraries: DGtal (DGtal) to represent 3D volumic data and for the computation of the trunk centerline, OpenCV (Bradski 2000) for the machine learning, GNU GSL (Gough 2009) for the linear regression and PCL (Rusu and Cousins 2011) for normal computation and clustering. The final objective of the project was to provide a Graphical tool for grading trunks or logs. The Graphic environment was provided by the platform Computree (<http://computree.onf.fr>), an open source software dedicated to processing Lidar data in forestry. The graphical interface which is currently available as a plugin of Computree and tested in a Linux operation system. We also proposed an alternative interface more rapid with no graphical interface in command line mode and allowing a batch process of multiple meshes.

## 3. Results

### 3.1. Defect Detection and Identification

Table 2 shows a comparison of the detection results thanks to the segmentation and classification algorithms with the reference data. The segmentation algorithm could detect most defects on the oak and beech trees barks even small defects such as picots and spheroblastes, the smallest diameter of detected defects being 0.5 cm. Overall, the segmentation algorithm detected 136 out of 141 defects in the meshes of six trunks. For each trunk, the missed defects were at most 2. As expected, the result of segmentation algorithm contained many false positives (false identification) resulting from miss-identified bark portions. The number of false positives was 410, but the classification algorithm was efficient to remove these false positives. The number of false positives remaining after the classification was 85. The classification, however, removed also 15 correct detected defects. The algorithm achieved a perfect result in case of Oak 334, which has only 8 burls; the algorithm removed all the false positives. On other trunks, the classification algorithm was less efficient with 7 and 8 false positives left in the other oak trees and from 22 to 26 on the three beech trees.

Table 2: Results of the segmentation (Seg.) and classification (Cla.) steps compared with the observed defects.

Tree name	Observed	Correct detected		False positive		Missed	
		Seg.	Cla.	Seg.	Cla.	Seg.	Cla.
Oak 334	8	8	8	9	0	0	0
Oak 411	25	24	20	147	8	1	5
Oak 569	24	23	22	79	7	1	2
Beech 101	31	30	26	58	22	1	5
Beech 102	29	29	25	70	22	0	4
Beech 104	24	22	20	47	26	2	4
Total	141	136	121	410	85	5	20

Figure 5 portrays the classification of the defects on the meshes of 6 trunks. In order to have a quantitative evaluation of the quality of the results, we compared the classification results with the reference data resulting from manual assessments on the field. The evaluation criteria were the  $F_1$  score and the average  $F_{\mu 1}$  (see Annex 1

for the definition of these criteria). Figure 6 shows the  $F_1$  score of each type in the step of classification. The average  $F_{\mu_1}$  was 0.8. There were obvious differences in the  $F_1$  score between the different defect types. The burl type had the best  $F_1$  score of 0.75, following by the branch type with  $F_1$  score of 0.67. The algorithm performed less well for branch scar and small defect with  $F_1$  scores of 0.54 and 0.52 respectively.

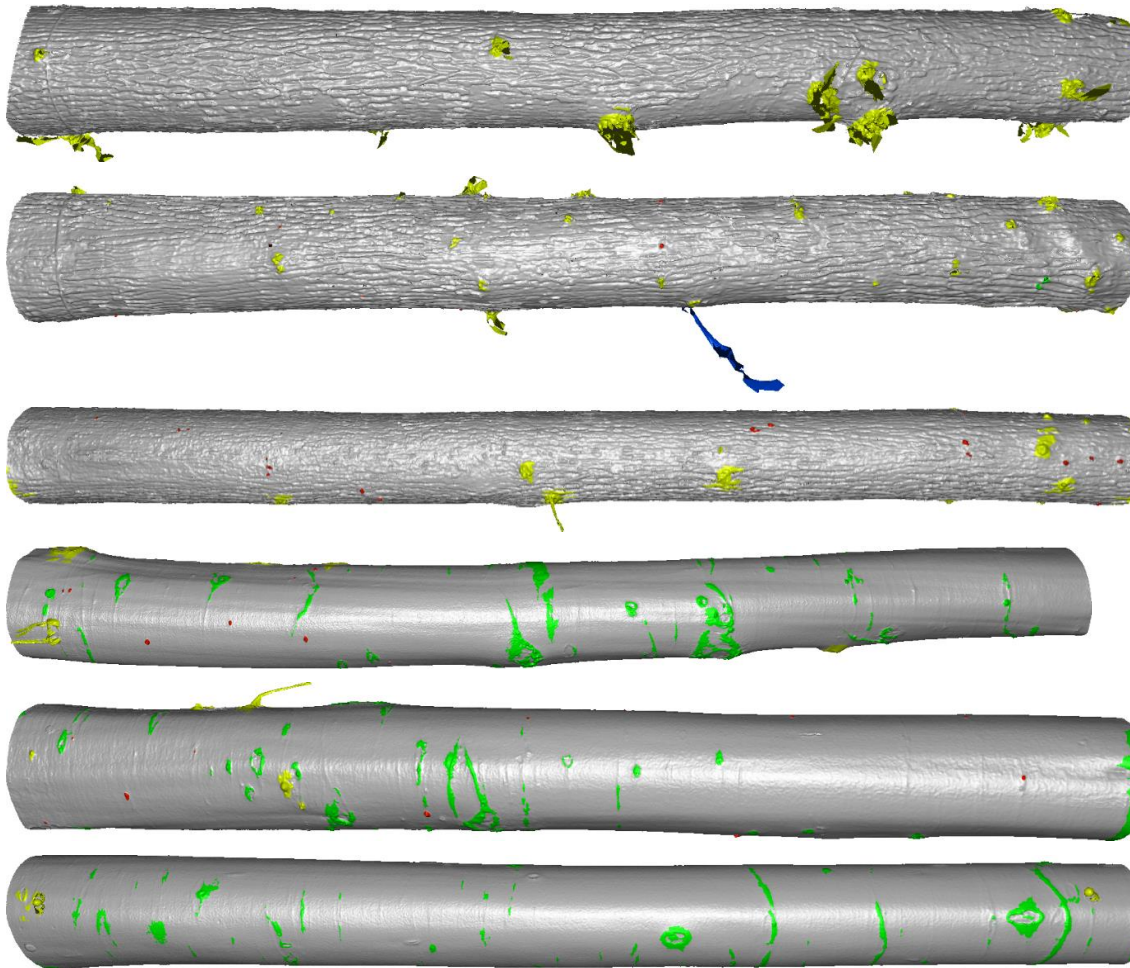


Figure 5: Visualization of the classification results in the meshes of three oak trees and three beech trees. The defects are highlighted according to their type. ■ is branch type including both sequential and epicormic branches. ■ is branch scar, ■ is Burl and ■ is small defect including bud cluster, spheroblaste and picot.

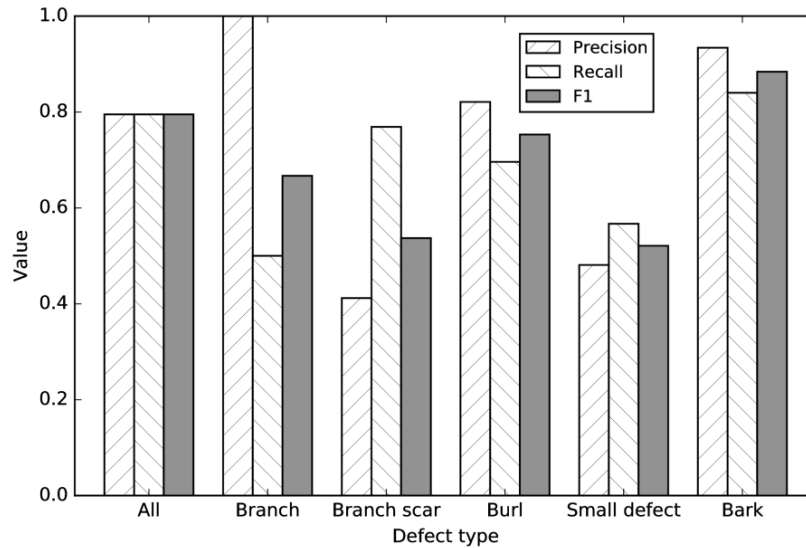


Figure 6: Precision, recall and  $F_1$  score of the different defect types. The Bark class was introduced because there were false detections of bark detected as defects by the segmentation algorithm. One of the purpose of the classification step was to remove these false detections.

### 3.2. Defect measurement

To evaluate the performance of our automatic measurement, we computed the difference between the automatic measurement and the manual measurement. The violin plots were used because they give more information about the distribution of the differences than a box plot. The violin plots in Figure 7 show the distribution of the differences of the width (left side) and height (right side) of branch scar, burl, small defects and also grouped into the overall category. It also shows the mustache width and height of branch scar. Overall, the algorithm achieved a good result with the median close to zero both for defect width (10.5 mm) and height (6.5 mm). The overall coefficient of variation for the absolute difference between the automatic measurement and the manual measurement of defect's width ( $cv_w$ ) and height ( $cv_h$ ) was 145% and 144% respectively.

#### *Branch scar*

The branch scar type had a relatively compact height difference and a small median of 2.9 mm. The difference of the branch scars width was, however, more disperse and with a larger median of 17.1 mm. The distributions of the plots were situated a little more on the positive side, but there is a considerable part of the plots situated below zero. The difference between the automatic measurement and the manual measurement of the mustache width had a considerable dispersion due to the weak relief of these parts in the defects making them very challenging to quantify. The coefficient of variation  $cv_w$  and  $cv_h$  were 134% and 147% respectively. The coefficient of variation of the mustache's width and height were 99% and 123% respectively.

#### *Burl*

The burl type has the most significant difference between the estimated and the observed value of width and height with medians of 36.5 mm and 23.9 mm respectively. The distributions were relatively compact but situated mostly over zero. The coefficient of variation  $cv_w$  and  $cv_h$  were 94% and 103% respectively.

#### *Small defect*

As it could be expected, the small defect type exhibits the most compact plots and has the smallest median width and height of 5.0 mm and 4.2 mm respectively. The distributions of the width difference and height difference are similar. The positive median and the large part of the plots of small defects are above zero suggesting that the



algorithm overestimated mostly the diameter of the small defects. The coefficient of variation  $cv_w$  and  $cv_h$  were 133% and 163% respectively.

Comparing the measurement error for different defect types, we can see that the absolute error of small defects is the smallest and is the largest for the branch scar. The coefficient of variation the difference between the manual and automatic measure of small defect type was, however, the largest due to its small mean (width mean of 18 mm and height mean of 13 mm).

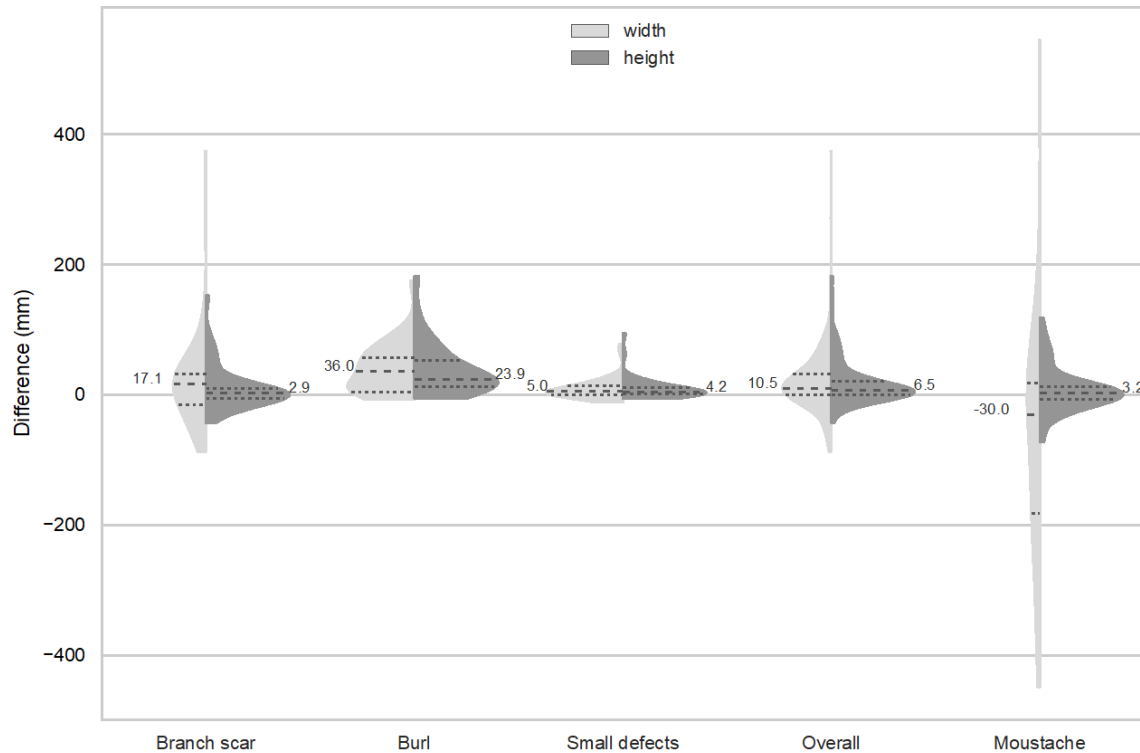


Figure 7: Violin plots of the differences between the estimated and the manual measurement of the defect dimensions for different defect types. For each plot, the dash lines show the quartiles and median with the latter accompanied by the number.

#### Branch diameter measurement

In our test database mainly focused on naturally pruned parts of the trunks, only two branches were measured, one in an oak tree and another on a beech tree. The results of the measurement are shown in Table 3, and the reconstruction of the two branches is illustrated in Figure 8.

Table 3: Branch diameter measurement result

Branch	Observed diameter (mm)	Estimated diameter (mm)
On Oak 411	8	10.29
On Beech 102	20	13.44

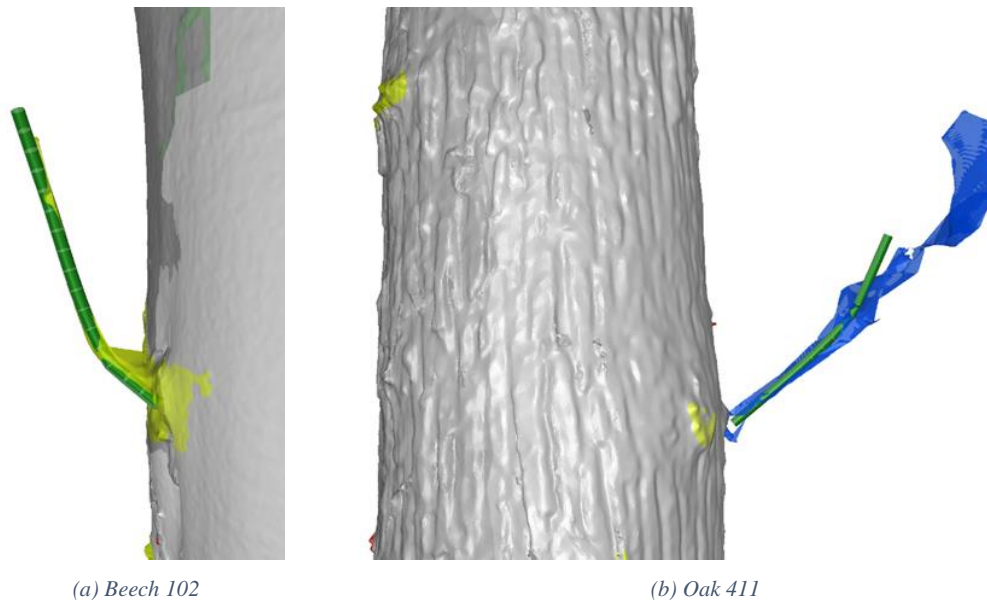


Figure 8: Illustration of branch measurement by computing its centerline. The reconstruction of the branch from the centreline and estimated diameter is shown in green.

#### Overall Species effect

The distribution of the differences is also different between the two species (Figure 9). The violin plot of the width of defects on oak is comparatively shorter which suggests that there is a good agreement between estimated and observed widths. The lower quantile is greater than 0 meaning that the algorithm overestimated the defect width for the oak trees. The violin plot of width on the beech trees is much taller which suggests that the agreement between estimated and observed widths was not as good. The algorithm tended to underestimate the large defects. The violin plots of the difference between the estimated and observed heights were shorter than the one of widths both on oak and beech. We also found that the coefficient of variation of the absolute difference between the estimated and the observed defect's dimensions on oak trees was smaller than on beech trees. The  $cv_w$  and  $cv_h$  of oaks were 107% and 111% respectively comparing to the  $cv_w$  and  $cv_h$  of beeches, which were 144% and 176% respectively. We can see that the absolute error of the measurement of the defect height was smaller than that of the defect width. The coefficient of variation of the height measurement was however smaller. That is because most defects had the height smaller than the width.

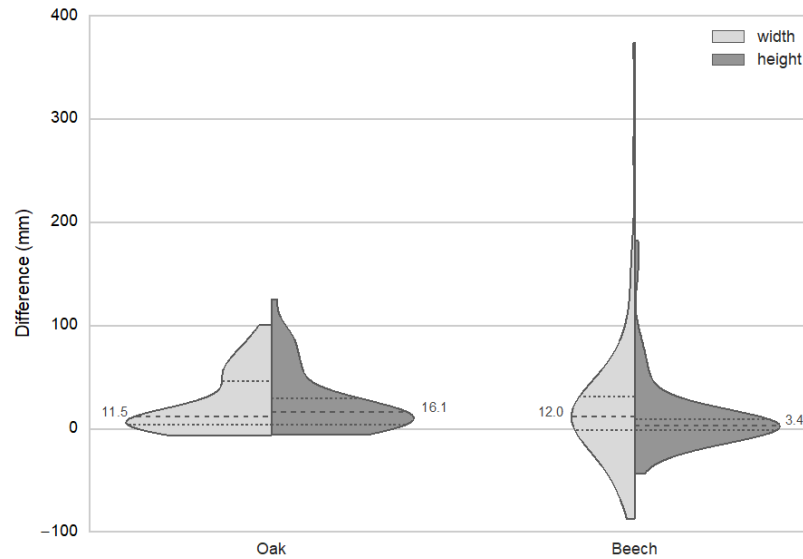


Figure 9: Difference between estimated and manual measurement of the defects dimensions on oak and beech

*Missing and false identifications*

The classification algorithm missed 20 out of 121 defects in total. Figure 10 shows the histograms of width and height of the missing defects. We can see that the algorithm missed only the small defects. The largest missing defect had a width of 45 mm. 70% of the missing defects had a width equal to or less than 20 mm. All missing defects had a height smaller than or equal to 25 mm.

The classification had 85 false identifications (false positives). Figure 11 shows the histogram of width and height of these false identifications. We can see that most false positives were relatively small of which 60% had a width that was less than 60 mm and 80 % had a height that was less than 30 mm.

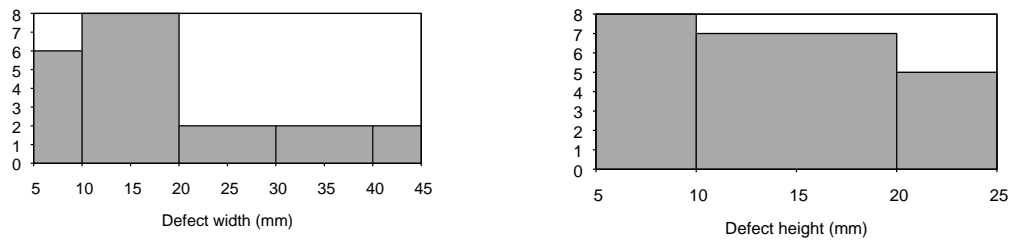


Figure 10: Defect width and height of missing defects

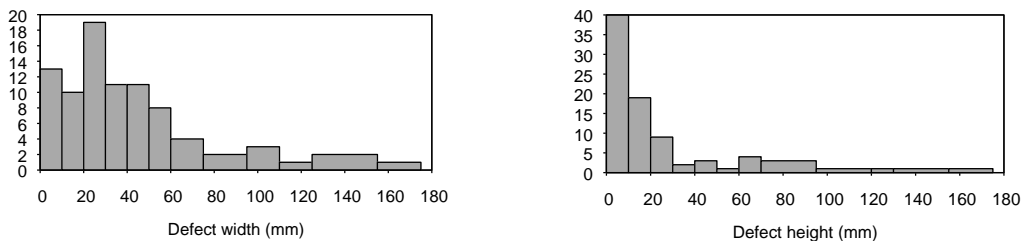


Figure 11: Defect width and height of false positives

### 3.3. Processing time

In this section, we show the running time of our program for the main process, which does not include the preprocessing time. The running time was recorded by the “time” program provided in GNU Linux operation system. The experiments were carried out on a laptop running GNU Linux with an Intel Core i7 processor

Figure 12 shows the running time of the program for 6 trunks and the main parameters, which could influence the running time. We can see that the running times depends on several factors because the program included several algorithms. The first factor was the number of points (vertex) in the meshes, as we can see in Figure 12 (a), if we fix others parameters, the general trend is that the execution time increased when the number of points increased. However, the factor that had the strongest impact on execution time was the voxel size. In Figure 12 (b), we can see that the execution time is almost inversely proportional to the square of the voxel size. As we increased the voxel size from 1 mm to 3 mm, the execution time decreased from 6 to 10 times. That was because using a larger voxel reduces data volume to process to compute the centerline and the patch query. The last main factor that had an impact on the execution time was the patch size. In our experiments, the execution time was proportional to the patch size. As shown in Figure 12 (b), when we doubled both width and height, the execution time was almost doubled. The raised question is what are the optimize parameters? In our experiments, the voxel size of 3 mm gave results that were as good as of the 1 mm voxel size but reduced the running time between 6 and 10 times. The choice of other parameters such as patch size and density depends on the characteristics of the tree and the defects on the trunk.

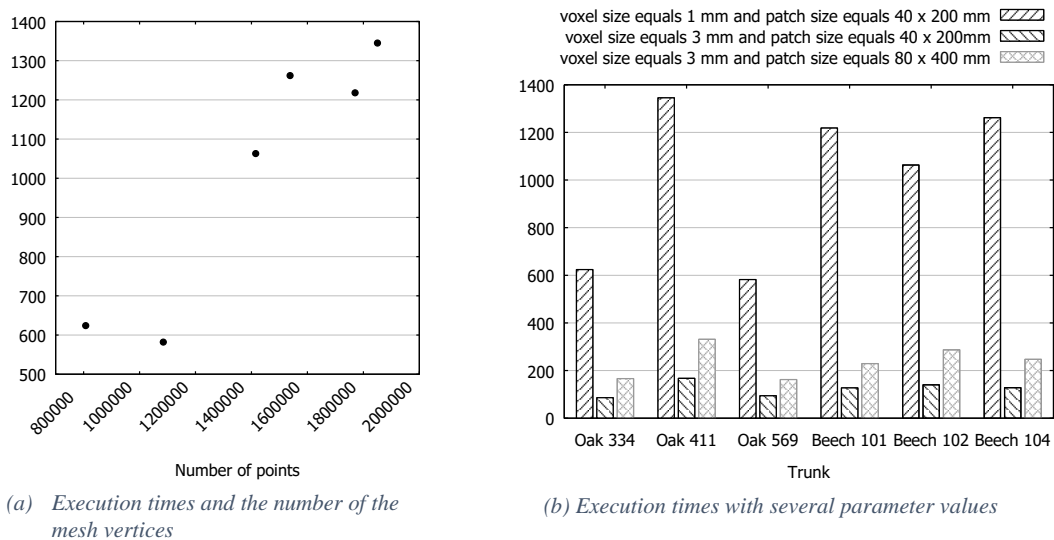


Figure 12: Summary of running times of the program on meshes using command line interface. The running times were counted for our 3 main steps. The time for preprocessing was not included.

## 4. Discussion

### 4.1. Defect detection and identification

The automatic detection of defects on the surface of the trunk using TLS data was a real challenge. Despite having at disposal an efficient method, many factors could impact the results such as species and age of the tree, global shape of the trunk, resolution, and quality of the scan. The species and age of the tree have a direct impact on the structure of the bark passing from smooth to furrowed as the tree ages or as we move from the upper to the bottom part of the trunk. Good quality and high-resolution data are required for a detection of the small defects.

In our experiment, for detecting the defect with a diameter of 0.5 cm a resolution of at least 20 points per cm<sup>2</sup> was required. From our results, we can say that our algorithms have an excellent performance in detecting even small defects. The detection outcomes of the segmentation algorithm were very good with only 5 missed defects out of 146 in total. There are two reasons allowing us to obtain the good detection rate. The first is that we have a very good and robust estimation of the trunk centerline. The second reason comes from our patch based approach to estimate the reference distance to the centerline for each individual point in the mesh of trunk. The efficiency of estimation allows detecting even slight modification of bark roughness. The smallest diameter of defects detected by our algorithm was 0.5 cm outperforming all previous studies with size ranging from 7.5 cm (Thomas et al. 2006) to 2 cm (Kretschmer et al. 2013). Moreover, our method takes all types of branching defects into account whereas the previous works focused only on branch scars.

The detection results, however, contained 410 false negatives for 141 manually detected defects. The classification algorithm allows to remove 325 of these false negatives. The most critical challenge in the classification of the defects would be to reduce the number of false positives. They indeed directly affect the result of the assessment method. As discussed in Nguyen et al. (2018), the performance of the classification might be improved by adding in the learning data set more defects having a large variability in shape. As shown in Figure 6, the precision of the classification of branch scars and small defects was low because of the false positives, in turn, due to the misclassification of bark portions as branch scars. It must be mentioned here that the developed method itself can help to build a most relevant learning database.

## 4.2. Quantification

Since the defect size was estimated depending on the defect type, we will discuss first the performance of the defect size estimation by type. Second, we will discuss the robustness of our algorithms to species. Third, we will discuss the potential utilization of our results to grading standing trees and roundwoods. Finally, we will discuss some limits of our approach and the TLS technology.

### 4.2.1 Comparison between defect types

#### *Branch scar*

Due to the particularity of branch scar shape, for inferring the knot size, we need the information about the scar seal width and height and the mustache width and height. The scar width and height were estimated with a median of 17.1 mm and 3.9 mm respectively. Compared to the burl, the branch scar had a better distribution of the differences between the automatic and manual measurements. Apart from that, two or more branch scars might have connected on a beech trunk. They have been recognized by our program as one branch scar. This phenomenon denoted as under-segmentation which led to a very high error in the estimation of the scar dimension that appeared in the violin plots with a difference greater than 300 mm in scar width. This kind of under-segmentation can be detected by a more detailed analysis of the shape of the detected defect and future developments can be imagined to solve this particularity. The performance of our algorithm on the measurement of scar height was similar to the result of Kretschmer et al. (2013) with 56% and 58% of measured scar height having a difference of less than 1 cm respectively.

#### *Burl*

In our experiments, burls were only present on oak trunks. Our algorithm often overestimated the diameter. 94% of both width and height were overestimated and medians of the error were 36 mm and 23.6 mm respectively. Our

sample oak trees had a furrowed bark resulting in small bark portion included in the detected defects (Figure 13). Consequently, the detected defects were often larger than the defects observed on the field, but this bias could be corrected if we can precise the reasons of its systematics bias.

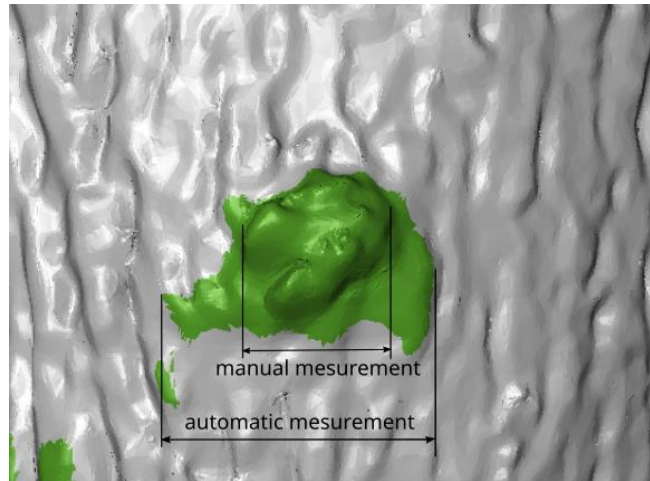


Figure 13: The difference between the manual and the automatic measurement of the burl diameter. The automatic measurement often leads to greater measures than the manual one.

#### *Small defect*

The dimensions of the small defects were also often overestimated with a rate of 73% for width and 76% for height. Since these defects have small dimensions, the relative measurement error usually high. For example, the coefficient of variation of the absolute error between the manual and automatic measurement for width and height were 133% and 163% respectively. However, in a wood quality point of view, the impact of these small defects is mostly by their number. The standard NF EN 1316-1 (AFNOR 1997) for example, treats a picot as a defect of diameter 5 mm.

#### *Branch diameter*

The estimation of the branch diameter was accurate, even though the result depends on the quality of the scan. The measurement of branch diameter was based on a reconstruction of the branch by a computation of the branch's main centerline. Figure 8 shows that the reconstruction was very robust. The branch on the Beech 102 (Figure 8 (a)) was very curved, and there was noise at the junction between the branch and the trunk. The branch reconstruction of the branch was very close to the real surface and shape of the branch. The estimated diameter of this branch was smaller than the manual measurement because the diameter of the branch decreases very fast from the intersection of the branch with the trunk. The scan of the branch on the Oak 411 (Figure 8 (b)) was very noisy. This comes from the fact that the branch was at the edge of the scan. The branch looked thus much larger and flatter than the real branch due to noisy contour. The algorithm returned however a very good estimated diameter of 10.29 mm compared to 8 mm from the manual measurement. The approach based on the branch's centerline accurate and did not require explicit parameters. It had fixed parameters such as accumulation radius, the skipped distance from the trunk bark and the length to estimate the diameter. In comparison with a cylinder-fitting-based method, we found that the centerline-based method was better. With the former, we need to provide an approximated radius of the cylinder close to the real radius of the branch which was not always measurable in a real scenario. Moreover, the fitted cylinders did not work well with a very curved branch as the beech branch showed in Figure 8 (b).

From the reconstruction of the branch from the centerline and estimated diameter, we can also estimate the insertion angle: the angle between the principal axis of the branch and this of the trunk. This information could be used to distinguish between a sequential branch and an epicormic branch since the former is more fastigiated. The incident angle could be also used to predict the knot shape and direction inside the trunk wood.

#### 4.2.2. Comparison between oak and beech

Our results showed that the method can work on both oak and beech which are the most common broadleaved species in Europe. There were, however, differences between the two species (Figure 9) due to the difference in the bark, global shape of the tree and defects types. The oak characteristics are a furrowed bark with often epicormic defects such as burl and picot. The branch scars on oak, however, are difficult to be detected from external characteristics of the bark since the bark may completely cover the scars. In the opposite, on the beech bark, the smooth bark reveals easily even the very small and weak relief defects. On the other hand, it is also very easy to get false detections. These false detections are more difficult to remove in comparison with the false detections occurring on oaks since they have a pattern very similar to old scars of rapidly pruned branches ridging the bark surface.

Beech bark often has anomalies due to *Nectria*. This disease caused by fungi and is one of the most common beech bark diseases at least in Northern France. The irregularities of the bark associated with the *Nectria* disease are mostly superficial and have little or no impact on wood quality (McCullough et al. 2005). These irregularities, however, were often miss-identified as branch scars by our algorithms. It would be possible to remove these false identifications in the classification step by adding a dedicated class for *Nectria*.

#### 4.2.3. Potential application to grading roundwoods and standing trees.

The results obtained in this study show clearly that the accuracy of defect measurements with TLS technology can be used to assess the wood quality of standing trees or roundwoods. It turns out that the use of the information delivered by our algorithm, such as position, type, and dimension of defects makes the automated grading of roundwoods or standing trees possible by standards such as NF EN 1316-1 (AFNOR 1997) grading oak or beech logs and Jourez et al. (2010) for grading broad leaf standing trees. However, we observed that grading a class A log is rather challenging whether by an automatic algorithm or by an operator. That is because this class tolerates no defects on a beech log or just a small defect such as picot or sound knot, which has a diameter of less than 15 mm. An error of the algorithm or an operator could lead to a miss evaluation from class A to B or vice versa. For a class B, these standards use the sum of the defects diameters for grading oak and beech logs. However, as shown above, the algorithm gave results often larger than the results from a manual evaluation. For the sum of diameters, the automatic measurement is about 50% larger than the manual measurement on oak, which might lead to an under evaluation of quality compared to a manual evaluation.

The method has still limits as the underestimation of branch scar edge due to the insufficiency of the relief and the under-segmentation of defects. The under-segmentation of defects may affect both the grading of standing trees and of roundwood. For example, an under-segmentation might count two defects as one, which results in a class B log graded as a class A. Thus, a further analysis is needed to detect under-segmentations and then correct them. These results need to be improved through the different possibilities already mentioned. Furthermore, the two methods mentioned are based on the evaluation of external defects, the correctness of the assessment needs thus to be validated by a measurement of knot inside wood, and the coupling between the outside and internal

characteristics of the defect can be improved by building learning database for classification making use of internal information coming from X-Ray CT scanning.

An important point in grading log quality is the identification between a sound knot and an unsound knot. To our knowledge, there was no research about this aspect of using TLS. For a completed automatic procedure of grading log quality using TLS, a study of the feasibility of the identification of these two-knot types is necessary.

The last aspect that we want to compare to a conventional procedure for grading standing trees such as of Jourez et al. (2010) is the processing time. As showed in the section “Results”, the required time for running our algorithms was small if we choose an optimized parameter, in particular, using the voxel size of 3 mm. The execution time for a nearly 2 million vertices mesh was from 2 to 3 minutes. However, if we count the time for acquiring and preprocessing the TLS data for a tree, this could be as long as 2 hours. The overall time required to process a tree by using TLS data was longer than the conventional procedure, which was about 30 minutes. Although the human vision is known to be better than the machine vision for an object recognition task, an operator, however, could make errors and inconsistent results depending on the emotion and working condition. A computer program, however, often gives a more consistent result. Moreover, with the advance of technology, we could incrementally improve the precision and reduce the processing time.

#### 4.2.4 Limitations of TLS

The TLS has several limitations such as the occlusion by branches, moss or parasite like ivy and the reduction of the spatial resolution in the high parts of tree trunks. A small cluster of moss can also look like a picot or a spheroblaste leading to an increase in the number of detected defects. This rationale does not consider that we only use geometrical information for processing the data. TLS data include at least reflectance data which can help also to classify the defects.

### 5. Conclusion

In this paper, we presented a method consisting of a suite of algorithms to measure the defects of the trunk surface from only the 3D information contained in TLS data. The method consists in three main steps. The first step was to detect the potential defect by a segmentation algorithm. The second step was to classify the detected defects into different defect types or bark. The focus of this paper is the last step which is to measure the position and dimension of defects. In our experiment made on 3 oak and 3 beech trees, the algorithms detected most defects even small defects such as picot and spheroblaste with a diameter as small as 0.5 cm. Due to the high variability of the defect shapes, the algorithms had an important number of false positives mostly due to the miss-classification of the bark portions as branch scars or small defects. Based on this suite, we proposed an automated approach to measure the characteristics of all kind of defects in agreement with the methods proposed in the grading rules of standing trees or roundwood. By comparing with the manual measurement, the automated measurement often resulted in more numerous and larger defects. But for a first attempt of comparison, the results are very promising considering that for characterizing the defects we used only 3D information compared to the assessment made by an experienced human operator. The method is fully compatible with the current methods used to estimate the quality of standing trees or roundwood, and we identified possible improvements to better estimate the inner quality. In regard to the ability to automatically detect and characterize different defect types, our method is a real improvement on the previous methods.



## References

- AFNOR (1997) EN 1316-1 - Hardwood round timber - Qualitative classification - Part 1: Oak and beech
- Bradski G (2000) The OpenCV Library. Dr Dobbs J Softw Tools
- Breiman L (2001) Random forests. *Mach Learn* 45:5–32
- Breiman L, Friedman J, Stone CJ, Olshen RA (1984) Classification and regression trees. CRC press
- Calders K, Newnham G, Burt A, et al (2015) Nondestructive estimates of above-ground biomass using terrestrial laser scanning. *Methods Ecol Evol* 6:198–208
- Carpenter RD, Jones M (1989) Defects in hardwood timber. US Department of Agriculture, Forest Service
- Colin F, Mechergui R, Dhôte J-F, Fontaine F (2010a) Epicormic ontogeny on *Quercus petraea* trunks and thinning effects quantified with the epicormic composition. *Ann For Sci* 67:813–813
- Colin F, Mothe F, Freyburger C, et al (2010b) Tracking rameal traces in sessile oak trunks with X-ray computer tomography: Biological bases, preliminary results and perspectives. *Trees - Struct Funct* 24:953–967. doi: 10.1007/s00468-010-0466-1
- Dassot M, Colin A, Santenoise P, et al (2012) Terrestrial laser scanning for measuring the solid wood volume, including branches, of adult standing trees in the forest environment. *Comput Electron Agric* 89:86–93
- DGtal DGtal: Digital Geometry tools and algorithms library
- Fonseca MA (2005) The measurement of roundwood: methodologies and conversion ratios. CABI
- Gough B (2009) GNU scientific library reference manual. Network Theory Ltd.
- Hackenberg J, Wassenberg M, Spiecker H, Sun D (2015) Non destructive method for biomass prediction combining TLS derived tree volume and wood density. *Forests* 6:1274–1300
- Henning JG, Radtke PJ (2006) Detailed stem measurements of standing trees from ground-based scanning lidar. *For Sci* 52:67–80
- Holopainen M, Vastaranta M, Kankare V, et al (2011) Biomass estimation of individual trees using stem and crown diameter TLS measurements. *ISPRS-Int Arch Photogramm Remote Sens Spat Inf Sci* 3812:91–95
- Hopkinson C, Chasmer L, Young-Pow C, Treitz P (2004) Assessing forest metrics with a ground-based scanning lidar. *Can J For Res* 34:573–583
- Hu M-K (1962) Visual pattern recognition by moment invariants. *IRE Trans Inf Theory* 8:179–187
- Jourez B, de Wauters P, Bienfait O (2010) Le classement des bois feuillus sur pied. *Silva Belg* 117:1–12
- Kankare V, Holopainen M, Vastaranta M, et al (2013) Individual tree biomass estimation using terrestrial laser scanning. *ISPRS J Photogramm Remote Sens* 75:64–75
- Kankare V, Joensuu M, Vauhkonen J, et al (2014) Estimation of the timber quality of scots pine with terrestrial laser scanning. *Forests* 5:1879–1895
- Kerautret B, Krähenbühl A, Debled-Rennesson I, Lachaud J-O (2016) Centerline detection on partial mesh scans by confidence vote in accumulation map. In: *Pattern Recognition (ICPR), 2016 23rd International Conference on. IEEE*, pp 1376–1381
- Kretschmer U, Kirchner N, Morhart C, Spiecker H (2013) A new approach to assessing tree stem quality characteristics using terrestrial laser scans. *Silva Fenn* 47:1–14. doi: 10.14214/sf.1071

- Litkey P, Liang X, Kaartinen H, et al (2008) Single-scan TLS methods for forest parameter retrieval. *Proc SilviLaser 2008*:8th
- McCullough DG, Heyd RL, O'Brien JG, Marquette M (2005) Biology and management of beech bark disease. *Mich State Univ Ext Bull E-2746*
- Meadows JS, Burkhardt E (2001) Epicormic branches affect lumber grade and value in willow oak. *South J Appl For* 25:136–141
- Mutanga O, Adam E, Cho MA (2012) High density biomass estimation for wetland vegetation using WorldView-2 imagery and random forest regression algorithm. *Int J Appl Earth Obs Geoinformation* 18:399–406
- Nguyen V-T, Constant T, Kerautret B, et al (2018) A machine-learning approach for classifying defects on tree trunk using terrestrial LiDAR (submitted)
- Nguyen V-T, Kerautret B, Debled-Rennesson I, et al (2016) Segmentation of defects on log surface from terrestrial lidar data. In: *Pattern Recognition (ICPR), 2016 23rd International Conference on*. IEEE, pp 3168–3173
- Othmani A, Voon LFLY, Stolz C, Piboule A (2013) Single tree species classification from Terrestrial Laser Scanning data for forest inventory. *Pattern Recognit Lett* 34:2144–2150
- Popescu SC, Wynne RH, Nelson RF (2002) Estimating plot-level tree heights with lidar: local filtering with a canopy-height based variable window size. *Comput Electron Agric* 37:71–95
- Pyörälä J, Kankare V, Vastaranta M, et al (2017) Comparison of terrestrial laser scanning and X-ray scanning in measuring Scots pine (*Pinus sylvestris* L.) branch structure. *Scand J For Res* 1–8
- Racko V (2013) Verify the accuracy of estimation the model between dimensional characteristics of branch scar and the location of the knot in the beech trunk. *Ann Wars Univ Life Sci-SGGW For Wood Technol* 84:
- Rusu RB, Cousins S (2011) 3D is here: Point Cloud Library (PCL). In: *IEEE International Conference on Robotics and Automation (ICRA)*. Shanghai, China
- Schütt C, Aschoff T, Winterhalder D, et al (2004) Approaches for recognition of wood quality of standing trees based on terrestrial laserscanner data. *Laser-Scanners For Landsc Assess Proc ISPRS Work Group VIII2 Freibg Ger Int Arch Photogramm Remote Sens Spat Inf Sci* 36:179–182
- Simonse M, Aschoff T, Spiecker H, Thies M (2003) Automatic determination of forest inventory parameters using terrestrial laserscanning. In: *Proceedings of the ScandLaser Scientific Workshop on Airborne Laser Scanning of Forests*. pp 252–258
- Stängle SM, Brüchert F, Kretschmer U, et al (2013) Clear wood content in standing trees predicted from branch scar measurements with terrestrial LiDAR and verified with X-ray computed tomography 1. *Can J For Res* 44:145–153
- Tague-Sutcliffe J (1992) The pragmatics of information retrieval experimentation, revisited. *Inf Process Manag* 28:467–490. doi: [https://doi.org/10.1016/0306-4573\(92\)90005-K](https://doi.org/10.1016/0306-4573(92)90005-K)
- Thies M, Spiecker H (2004) Evaluation and future prospects of terrestrial laser scanning for standardized forest inventories. *Forest* 2:1
- Thomas L, Shaffer CA, Mili L, Thomas E (2006) Automated detection of severe surface defects on barked hardwood logs. *Forest*
- Thomas L, Thomas E, others (2010) A graphical automated detection system to locate hardwood log surface defects using high-resolution three-dimensional laser scan data. In: *Proceedings of the 17th Central Hardwood Forest Conference*. pp 5–7
- Thomas RE (2009) Modeling the relationships among internal defect features and external appalachian hardwood log defect indicators. *Silva Fenn* 43:447–456

- Wang D, Hollaus M, Puttonen E, Pfeifer N (2016) Fast and robust stem reconstruction in complex environments using terrestrial laser scanning. *Int Arch Photogramm Remote Sens Spat Inf Sci* 41:
- Yao T, Yang X, Zhao F, et al (2011) Measuring forest structure and biomass in New England forest stands using Echidna ground-based lidar. *Remote Sens Environ* 115:2965–2974
- You L, Tang S, Song X, et al (2016) Precise measurement of stem diameter by simulating the path of diameter tape from terrestrial laser scanning data. *Remote Sens* 8:717

## Annex 1

*Evaluation criteria for detection and classification.*

The criteria used to evaluate the performance of the detection and classification were the  $F_1$  score (F-measure), which is the harmonic mean of the precision (P) and the recall (R). The  $F_1$  score of the classification result for each type is computed from the precision and the recall as follows:

$$F_1 = \frac{2PR}{P + R}$$

The “precision is defined as the number of correctly classified positive examples divided by the number of examples labelled by the system as positive:

$$P = \frac{TP}{TP + FP}$$

The “recall” is the number of correctly classified positive examples divided by the number of positive examples in the data:

$$R = \frac{TP}{TP + FN}$$

The overall F-measure ( $F_{\mu 1}$ ) is computed from the averaged precision ( $P_{\mu}$ ) and recall ( $R_{\mu}$ ) (Tague-Sutcliffe 1992) by the following equations:

$$F_{\mu 1} = \frac{2P_{\mu}R_{\mu}}{P_{\mu} + R_{\mu}}$$

$$P_{\mu} = \frac{\sum_{i=1}^l TP_i}{\sum_{i=1}^l (TP_i + FP_i)}$$

$$R_{\mu} = \frac{\sum_{i=1}^l TP_i}{\sum_{i=1}^l (TP_i + FN_i)}$$

where  $TP_i$  and  $FP_i$  are the true and false positive of class  $i$  (defect type  $i$ ) respectively.



## Chapitre 5

# Discussion et perspectives

### 5.1 Synthèse de la méthode et des résultats principaux

L'objectif principal de ce travail était de développer une méthode participant à l'évaluation de la qualité d'arbres sur pied à partir de données tridimensionnelles en y apportant la contribution des défauts externes du bois. Nous avons décidé de proposer une approche en trois étapes : première étape : détection des défauts; deuxième étape : identification des types des défauts; troisième étape : caractérisation des défauts.

La première étape est basée sur un calcul de la ligne centrale et l'utilisation d'un voisinage du point allongé dans la direction longitudinale. Ainsi à chaque point, la distance de référence à la ligne centrale est calculée à partir de la distribution des distances à la ligne centrale de tous les points du voisinage. Un seuillage automatique (Rosin, 2001) est utilisé sur la distribution des écarts entre la distance du point et la distance de référence pour obtenir les écarts maximaux supposés correspondre à des points appartenant à des défauts.

La deuxième étape consiste à mettre en œuvre une approche d'apprentissage automatique se basant sur les forêts aléatoires de Breiman (2001), i.e. Random Forests, pour classer les défauts détectés par la première étape. Le vecteur des caractéristiques comprend l'espèce, les moments invariants de Hu (Hu, 1962), les dimensions du défaut, les valeurs propres en lien avec les axes principaux du défaut, et la moyenne et l'écart-type de la différence entre les distances à la ligne centrale et les distances de référence à la ligne centrale.

La dernière étape consiste à estimer la position et les dimensions des défauts. La méthode d'estimation dépend du type de défauts. Par exemple, le diamètre de branche est estimé par la méthode de la ligne centrale suivie de la recherche du mode de la distribution des distances des points à la ligne centrale. Pour une cicatrice de branche, une méthode basée sur l'analyse des profils verticaux du relief est proposée pour estimer les dimensions de la cicatrice et de la moustache.

Nous avons utilisé deux méthodes pour évaluer la performance de la détection. La première méthode jugeait la performance au niveau des points de la surface du maillage. Dans une expérimentation menée sur 10 billons de 7 espèces, le résultat de l'algorithme a été comparé avec une référence faite par des experts. La mesure moyenne de  $F_1$  était de 0,71 et de 0,76 pour la version améliorée. La deuxième méthode estimait la performance au niveau de l'ensemble des défauts. Dans l'expérimentation menée sur 26 troncs de 6 espèces, notre algorithme a montré une excellente performance dans la détection même de petits défauts. L'algorithme a détecté 180 défauts sur 185 au total.

Deux raisons expliquent ce bon taux de détection. La première est que nous avons une méthode très efficace et robuste d'estimation de la ligne centrale du tronc. La seconde est que notre approche estime la distance de référence à la ligne centrale pour chaque point individuel dans le maillage du tronc en analysant la distribution spatiale des points voisins sur un patch de forme allongée. L'efficacité de l'estimation permet de détecter même une légère modification de la rugosité de l'écorce. Le plus petit diamètre de défauts détecté par notre algorithme est de 0,5 cm. Notre méthode est donc plus performante que toutes les méthodes proposées avec une taille allant de 7,5 cm (Thomas et al., 2007) à 2 cm (Kretschmer et al., 2013). De plus, notre méthode prend en compte tous les types de défauts de ramification alors que les travaux précédents se concentraient uniquement sur les cicatrices de branches. Cependant cette recherche d'exhaustivité dans la détection, entraîne un nombre considérable de faux positifs : 762 dans cette expérimentation.

Néanmoins, l'étape suivante d'identification des défauts joue son rôle et réduit considérablement le nombre des faux positifs, qui descend de 762 à 111, mais elle affecte aussi le nombre des défauts correctement identifiés qui passe à 163 sur 185 au total. La valeur moyenne de la mesure  $F_1$  est alors de 0,84.

L'étape ultime de caractérisation des défauts a été expérimentée sur 3 troncs de chêne et 3 troncs de hêtre. Le résultat de l'estimation des dimensions de défauts est apparu prometteur avec la médiane des différences entre les diamètres horizontaux estimés et observés de 10,5 mm, et celle des diamètres verticaux de 6,5 mm. Cependant, les dimensions sont souvent surestimées et la distribution de la différence entre les dimensions estimées et observées est assez dispersée.

Les informations des défauts produites par la méthode telles que leur type, leur position et leurs dimensions peuvent être utilisées pour évaluer la qualité de l'arbre sur pied. La méthode peut être encore améliorée, la suite de ce chapitre discute des améliorations envisageables et aussi des perspectives.

Notre méthode est reproductible avec le code de la partie segmentation des défauts est disponible en ligne et aussi une démonstration en ligne. Le code des autres parties sera aussi disponible en ligne.

## 5.2 Détection des défauts externes à partir des données TLS

### 5.2.1 Résolution spatiale des données TLS et dimension des défauts détectables

Dans notre expérience, pour détecter un défaut avec un diamètre de 5 mm, une résolution d'au moins 25 points par  $\text{cm}^2$  est nécessaire. Pour atteindre cette qualité de données, il est nécessaire de positionner le scanner entre 3 m et 5 m de distance au tronc et utiliser au moins la moitié de la résolution maximale. La résolution angulaire maximale (pas angulaire) des scanners actuels est souvent  $0.009^\circ$  (Faro) et  $0.018^\circ$  (Leica) et conduit à une distance d'environ 0,5 mm (Faro) et 1 mm (Leica) entre deux points sur une surface à 3 m. A noter que si on utilise la résolution maximale et une distance courte entre le scanner et le tronc, le nuage de points est trop dense et une partie des données est superflue. La surabondance de données cause une exigence inutile sur la puissance de la machine et le temps de traitement. Si le scanner est trop proche du tronc, il y a aussi deux effets indésirables. Le premier est l'apparition d'une différence de densité entre la partie basse et la partie haute du tronc non négligeable. Par exemple, avec un scanner qui a une résolution angulaire maximale de  $0.009^\circ$ , et si le scanner est à 2 m du tronc, la densité de données est divisée par 5 quand on passe d'une hauteur sur le tronc de 1 m à 10 m. Le deuxième est la perte de précision dans la partie haute du tronc quand le scanner utilise la technique du décalage de phase à cause de l'angle d'incidence trop grand (Soudarissanane et al., 2009). La distance conseillée se situe donc entre 3 m et 5 m.

La détection des plus petits défauts est possible avec des données de haute résolution. Les petits défauts sont par exemple le bourgeon isolé qui a un diamètre d'environ 3 mm et un relief très faible. Dans les travaux réalisés qui sont ciblés sur l'aspect qualitatif du bois, nous n'avons donc pas traité ce type de défauts car il n'a aucun impact sur la qualité du bois. Cependant, la détection des défauts de ce type a un intérêt pour prédire la qualité future de l'arbre car les bourgeons sont le réservoir d'une partie des défauts potentiels que nous visons. Dans ce cas, la détection des bourgeons isolés est faisable sur les jeunes arbres qui présentent une écorce encore assez lisse, et en utilisant des données TLS à très haute densité. Toutefois, des difficultés envisagées pour la détection des bourgeons isolés sont (1) le bruit dans le nuage de points, lié aux échos multiples, qui est plus important sur de petits diamètres, (2) un très grand niveau de détail entraînant plus de faux positifs et (3) la résolution et la qualité de données en partie haute du tronc qui pourraient être insuffisantes comme évoqué précédemment. En se basant sur la résolution des scanners actuels (Faro), à 5 m du tronc, pour détecter un objet ayant un diamètre de 3 mm nous disposons d'une résolution de 0,8 mm.



## 5.2.2 Jeux de données et temps de traitement

Le temps de traitement total pour une approche basée sur les données LiDAR comprend le temps d'acquisition, le temps de prétraitement des données (fusion, isolation du tronc, lissage et maillage) et le temps de traitement principal (détection, classification et caractérisation). Nous avons observé que le temps de traitement contribue très peu au temps total. Par exemple, le temps du traitement principal pour un maillage dont le nombre de points est environ 2 millions est de 2 à 3 minutes avec un paramétrage assez robuste dont une taille de voxel de 3 mm et des largeur et hauteur de patch de 50 mm et 200 mm respectivement. Le temps d'acquisition et de prétraitement est environ 2 heures pour un arbre. En comparant avec le temps nécessaire pour une procédure conventionnelle (environ 30 minutes pour relever les défauts d'un arbre), le temps total requis par l'approche utilisant le TLS est encore trop grand surtout du fait du temps d'acquisition qui constitue le frein principal à l'essor de cette technologie dans le milieu forestier.

Dans les opérations de classement traditionnel, un opérateur doit être un expert et peut aussi commettre des erreurs et obtenir des résultats variables dépendant de la fatigue et des conditions de travail. Si la question de l'acquisition était résolue, un programme informatique, pourrait garantir une certaine expertise, devrait rendre plus facile la mesure de certaines variables et constituer une aide à la décision.

Avec les progrès de la technologie, certaines limitations sont réduites en temps. Ainsi avec l'avancée de la technologie, certaines limitations en temps ont déjà été réduites. Par exemple, comparons le temps d'acquisition et la précision de deux scanners utilisés à savoir le FARO ® LS 840 et le FARO ® Focus 3D X130 HDR. Le premier est sorti en 2005 ; il offre une résolution angulaire de  $0.009^\circ$  et l'erreur est de 3 mm. La vitesse maximale de la mesure est de 120.000 points par seconde. Le second est un appareil qui peut produire des scans avec la même résolution angulaire mais avec une erreur plus petite (de 2 mm) et une vitesse de mesure maximale 8 fois plus grande que le premier.

Plus généralement en 10 ans on a assisté à des progrès énormes, d'une part, dans la miniaturisation des appareils en passant de 20 kg à 1 kg même si dans ce cas les performances sont un peu limitées pour certaines applications envisagées, et d'autre part dans la réduction de leur prix, divisé par 3, voire par 4. Comparativement, leurs performances évoluent moins vite, mais les scanners 3D deviennent de plus en plus abordables financièrement et la technologie se généralise notamment grâce aux développements des véhicules autonomes.

## 5.2.3 Occlusion

L'occlusion est une des principales limitations de la technologie TLS qui peut réduire la performance de la détection. L'occlusion arrive s'il y a des objets qui créent des ombres sur le tronc. Nous

observons deux types d'occlusion. Le premier est causé par des branches, des feuilles ou d'autres arbres qui cachent une partie du tronc. Ce type d'occlusion empêche la détection des défauts. Une occlusion de ce type peut être réduite en acquérant les données de différents points de vue sachant que cela peut augmenter le temps nécessaire pour acquérir et traiter des données. Le deuxième est causé par les lierres et les mousses sur le tronc qui peuvent masquer les défauts ou créer de fausses détections. Par exemple, un petit amas de mousse sur l'écorce peut être reconnu comme un picot ou un sphérobaste, qui sont difficiles à distinguer dans un nuage de points puisqu'on n'utilise que les informations des coordonnées 3D du tronc. Une piste d'amélioration est d'utiliser la valeur d'intensité de points comme une source supplémentaire d'information. Cependant, cela rajoute de la complexité telle que la conservation de l'information de l'intensité pendant le prétraitement des données et de plus la précision de l'intensité des points dépend de la lumière ambiante (Voegtle et Wakaluk, 2009) ce qui diminue la fiabilité. Les mêmes commentaires s'appliquent sur l'ajout d'une information colorimétrique par fusion du nuage et de photographies couleurs, avec le temps de réalisation des photographies en plus.

### **5.3 Techniques d'identification de défauts : comment améliorer ?**

Nous avons observé que les formes de défauts et de l'écorce varient selon les espèces, l'âge, les conditions de développement de l'arbre. Cette variabilité exige une grande base d'apprentissage pour avoir la meilleure performance. Comme discuté dans la Chapitre 3, la performance de la classification peut être améliorée en ajoutant des exemples de défauts dans la base d'apprentissage. Il faut noter que ces exemples doivent couvrir la variabilité de formes de défauts, l'effectif de défauts seul n'indiquant pas une meilleure performance. Une autre piste pour améliorer la performance de l'apprentissage automatique est d'utiliser plusieurs bases d'apprentissage, une pour chaque espèce. Notons simplement que la reconnaissance d'espèces n'est pas complètement aboutie (Othmani et al., 2013). La séparation de la base d'apprentissage conduirait à une meilleure adaptation aux formes des défauts pour les deux raisons suivantes. (1) Une espèce possède un ensemble de types de défauts qui peut être différent de celui d'autres espèces. Par exemple, le chêne présente les types de défauts suivants : le gourmand, l'amas de bourgeons, le broussin et le picot. Le hêtre n'a pas de picot mais a des sphérobastes. L'épicéa n'a pas d'épicormiques. (2) Les formes des défauts d'un même type mais sur des espèces différentes peuvent être également différentes.

L'ajout des défauts supplémentaires aux bases d'apprentissage a néanmoins un coût important pour les raisons suivantes. 1) L'acquisition des données inclue le temps d'acquisition des données et le temps pour relever des défauts ; 2) le prétraitement des données ; 3) la segmentation des défauts, soit automatiquement par notre programme, soit manuellement sachant que la segmentation manuelle est très délicate notamment sur les données à haute densité ; 4) la vérification du type des défauts par rapport aux données relevées sur le terrain.

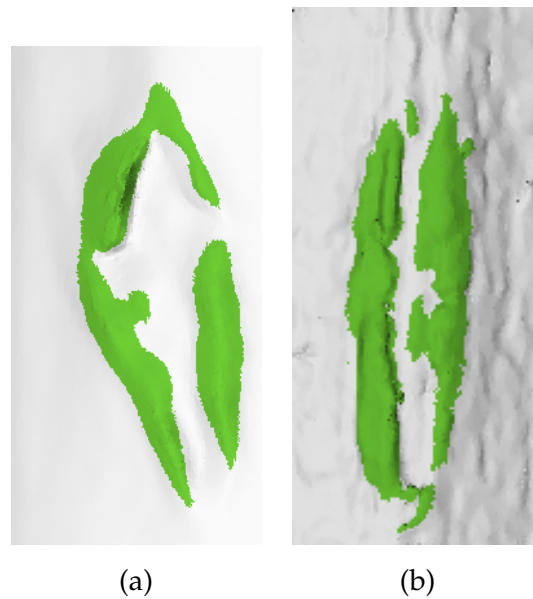
Une autre amélioration possible pourrait être une utilisation d'une méthode d'apprentissage plus avancée que les Random Forests telle que l'apprentissage approfondi, i.e. deep learning (LeCun et al., 2015). L'intérêt de cette méthode a été confirmé dans plusieurs études pour sa performance dans la reconnaissance d'objets et particulier pour les données LiDAR (Zhu et al., 2017; Maturana et Scherer, 2015; Guan et al., 2015).

## 5.4 Influence de l'espèce en dehors des types de défauts

Malgré la mise à disposition d'une méthode efficace, de nombreux facteurs peuvent avoir un impact sur les résultats tels que l'espèce et l'âge de l'arbre, la forme globale du tronc, la résolution et la qualité de données. L'espèce et l'âge de l'arbre ont un impact direct sur la structure et l'épaisseur de l'écorce. Certaines espèces ont une écorce lisse pendant toute leur vie; c'est le cas du hêtre. D'autres espèces ont une écorce qui peut passer de lisse à crevassée au fur et à mesure que l'arbre vieillit et la profondeur de crevasse diminue sur la partie haute du tronc. Avec l'utilisation des patches, notre méthode est à même de s'adapter à l'évolution de la rugosité de l'écorce avec la hauteur ou entre espèces pour détecter les défauts, néanmoins la rugosité de l'écorce est un d'attribut important d'une espèce et de la signature de ses défauts. Dans cette thèse, nous avons testé plusieurs billons ou troncs d'espèces différentes. Parmi ces espèces testées, nous pouvons classer les structures de l'écorce en quelques catégories. (1) Écorce lisse : hêtre, merisier et chêne rouge jeunes. (2) Écorces crevassées : chêne âgé. (3) Écorces fissurées : Orme. (4) Écorces se détachant par écailles : épicéa. (5) Écorces se détachant par plaques : mélèze. Ces caractéristiques des écorces ont une influence sur les algorithmes, notamment il faut enrichir les données de la classe "écorce" de la base d'apprentissage pour couvrir la variété de la structure des écorces. La rugosité de l'écorce est aussi un défi pour détecter les petits défauts tels que l'amas de bourgeons et les bourgeons isolés.

## 5.5 Prise en compte des défauts non liés à la ramification

En plus des défauts liés à la ramification, la surface du tronc peut présenter d'autres types de défaut : on peut citer les blessures d'exploitation, les fentes causées par le séchage, les gélivures, des maladies causées par les champignons, les blessures causées par les animaux. L'impact de ces défauts sur la qualité est très variable. Parmi ces défauts, la maladie du T chez le hêtre, les fentes et les gélivures sont très pénalisantes selon la norme AFNOR (2012a). Les maladies causées par les champignons telles que le *Nectria* ont un impact moins important sur la qualité du bois. En revanche, elles peuvent avoir un impact significatif pour la survie de l'arbre (McCullough et al., 2005). La détection de ces défauts est donc nécessaire soit du point de vue de la qualité du bois soit de la gestion de la forêt. La détection et l'identification de ces types de défaut est possible car ils



**Figure 5.1** Une maladie du T sur un hêtre (a) et une fente sur un chêne détectées par l’algorithme proposée.

ont des caractéristiques particulières. Par exemple, un défaut causé par la maladie du T sur le hêtre a une forme elliptique ; une fente a une forme crevassée verticale avec des reliefs sur les bords. En général, la méthode actuelle peut détecter ces types de défauts (Figure 5.1). Cependant, cela nécessite des données supplémentaires dans la base d’apprentissage pour décrire ces défauts et les distinguer. En particulier, il est nécessaire de créer une classe de défaut pour chaque type de défaut dans la base de d’apprentissage. Ensuite, il faut entraîner les Random Forests avec la nouvelle base d’apprentissage incluant des nouvelles données de ces classes par la procédure évoquée dans la section 5.3.

## 5.6 Une analyse plus fine des défauts traités

### 5.6.1 Classification des branches séquentielles et des branches épïcormiques

Dans le cadre de nos résultats, nous pouvons identifier les branches et estimer certains de leurs paramètres. Les paramètres sont le diamètre et l’inclinaison à leur insertion. Ces paramètres peuvent permettre de distinguer une branche séquentielle et une branche épïcormique car sur un arbre, les branches séquentielles ont un diamètre plus grand et un axe plus redressé que les branches épïcormiques. A part ces deux caractéristiques, la forme de la cicatrice peut être utilisée comme une source supplémentaire pour distinguer une branche séquentielle et une branche épïcormique. En effet, les cicatrices de branche séquentielle sont en forme de moustache de chinois alors que celles de branches épïcormiques sont circulaires (Colin et al. 2018, à paraître) (Figure 5.2). De plus,



(a) Branche séquentielle

(b) branche épïcormique

**Figure 5.2** Différence en forme de la cicatrice d'une branche séquentielle (a) qui a une forme de "moustache de Chinois" et d'une branche épïcormique (b) qui a une forme ronde.

une analyse des moustaches de la branche peut fournir des informations supplémentaires pour une analyse plus détaillée de la trajectoire de défaut à l'intérieur du bois : inclinaison, épaisseur de recouvrement du nœud (Wernsdörfer et al., 2005).

### 5.6.2 Classification des nœuds sains et des nœuds pourris

Un point que nous n'avons pas encore abordé est l'identification des nœuds pourris. Ce type de nœud est sévèrement pénalisant dans la procédure de l'estimation de la qualité du bois parce que l'impact sur la qualité mécanique et esthétique est plus important que pour un nœud sain du même diamètre. Il correspond à une partie de branche morte progressivement englobée dans le tronc et qui a eu le temps de se dégrader avant d'être recouverte. Cette différence a sans doute un impact sur la morphologie de la cicatrice mais il faudrait le mettre en évidence par une étude spécifique qui pourrait aussi se baser sur de la tomographie à rayons X. Ensuite, un objectif serait de séparer la classe cicatrices de branches en deux classes. La première classe comporterait les cicatrices correspondant aux nœuds sains et la deuxième classe, les cicatrices correspondant aux nœuds pourris. Le classifieur serait entraîné par des nouvelles données contenant ces deux classes.

## 5.7 Applications potentielles

### 5.7.1 Un outil pour classer des grumes ou des arbres sur pied

La version actuelle du logiciel ne comprend que les processus principaux tels que le calcul de la ligne centrale, la détection de défauts, la quantification de défauts. Les étapes de prétraitement doivent être réalisées par d'autres logiciels tel que Graphite (<http://alice.loria.fr/software/graphite>) pour faire un lissage et un maillage. Un des objectifs pour un futur travail serait d'intégrer toutes les étapes dans une chaîne de traitement complète qui proposerait aux utilisateurs d'entrer un nuage de points d'un tronc ou d'un billon et de sortir des résultats. Une piste envisageable est d'intégrer la bibliothèque Geogram (<http://alice.loria.fr/software/geogram>) dans la plateforme Computree. Cette bibliothèque offre des implémentations des algorithmes pour faire les lissages et les maillages utilisés par le logiciel Graphite.

Comme présentée dans Chapitre 4, la détection de défauts externes sur le tronc d'arbre présente des perspectives intéressantes pour améliorer les procédures conventionnelles d'évaluation de la qualité du bois selon les normes en vigueur (AFNOR, 2012a; Jourez et al., 2010). Cependant, une version étendue de l'application actuelle doit être envisagée qui prendra en compte non seulement les défauts mais aussi les attributs globaux exigés par ces normes tels que la courbure et les diamètres du tronc. L'application devra quantifier les autres types de défauts externes évoqués précédemment tels que les fentes, les gélivures ou encore la maladie du T. De plus, les attributs de la grume ou billon tels que leurs diamètres, leurs volumes et leurs courbures pourraient être déduits à partir des données TLS comme le font déjà certaines étapes du plugin Silva de Computree. Les courbures du tronc peuvent être déduites directement de la ligne centrale alors que les diamètres ou les volumes peuvent être calculés à partir des distances à la ligne centrale des points dans le nuage de points ou par une reconstruction de la surface du tronc (Morel et al., 2018).

Comme évoquée dans la section 5.3, la construction de la base de données d'apprentissage prendra un temps considérable. Du point de vue de l'architecture de l'application, l'alimentation de la base de données pourrait être progressive. En effet, la base pourrait être stockée sur un serveur ce qui permettra à l'utilisateur d'envoyer les défauts confirmés pour les ajouter à la base d'apprentissage.

### 5.7.2 Adaptation pour les bois ronds

Bien que la segmentation ait été testée avec des arbres sur pied et des billons, la quantification des défauts a été testée uniquement sur des arbres sur pied. Une adaptation de l'application serait nécessaire pour maximiser les résultats sur les bois ronds. En plus du fait que l'information n'est pas accessible sur toute la périphérie, ces billons subissent de nombreuses altérations du

fait de l'abattage, de l'élagage, et des nombreuses manipulations. Elles modifient complètement l'information sur les branches qui sont souvent coupées à raz du tronc et rendent difficile l'estimation de leur diamètre à partir de données 3D. De même l'apparence de l'écorce est fortement altérée par des arrachements et la présence de terre qui risque de générer d'autres type de faux positifs. Il faut aussi prévoir une procédure pour intégrer les informations disponibles sur les extrémités des billons (fentes notamment). D'autres projets comme l'ANR Treetrace (2017-2020 ANR-17-CE10-0016 ) s'intéressent à cette question.

## **5.8 Mise en relation avec les données de tomographie à rayons X**

La prochaine étape à développer sera de déduire des paramètres internes de défauts à partir des informations sur les défauts externes produites par le logiciel. Ainsi des informations de défaut externe telles que la position, le type et les dimensions permettraient de reconstruire les structures (traces) de ramification dans le bois pour fournir une information un peu similaire à celle délivrée par un scanner à rayons X

Avant de faire cela, une étape de construction des relations entre la morphologie externe des défauts et leur morphologie interne est à réaliser en les comparant avec des informations de référence. Deux approches sont possibles pour cette construction. La première est une approche destructive en coupant le bois pour évaluer les traces de ramification dans le bois (Thomas, 2009). La deuxième est l'utilisation de la reconstruction de traces de ramification ou des nœuds sur les images 3D venant de la tomographie à rayons X (Colin et al., 2010b). En effet, pour obtenir une estimation précise de la qualité du bois, il faut non seulement connaître la taille et l'état des défauts mais aussi connaître leur trajectoire dans le bois. Comme la tomographie à rayons X permet de reconstruire les trajectoires d'un grand nombre de défauts dans le bois, notamment liés à la ramification, cette technologie est une bonne référence pour construire les relations entre les caractéristiques des défauts externes et celles des défauts internes. Une piste pour améliorer la méthode actuelle est de reconstruire les défauts internes à partir des informations déduites de défauts externes telles que position, dimensions, ratio entre la largeur et hauteur, moustaches (Torkaman et al., 2018; Osborne et Maguire, 2015; Racko, 2013; Stängle et al., 2013; Duchateau et al., 2013). Néanmoins, une difficulté à prévoir pour cette construction est la fusion dans un repère commun des deux types de données du fait d'une densité d'information différente en surface, et plus spécialement dans le cas où un billon ne présente pas de singularités pouvant servir de base au recalage.





# Conclusion

Dans le contexte forestier, le LiDAR terrestre est utilisé à des fins scientifiques pour caractériser la distribution spatiale des volumes de bois dans un arbre ou dans une placette (Dassot et al., 2012; Kwak et al., 2014; Raunonen et al., 2013; Côté et al., 2011), ou encore l'interception de la lumière par le feuillage (Moorthy et al., 2008). Sa capacité à apporter des informations originales, tridimensionnelles et souvent difficiles à mesurer, comme par exemple les courbures (Liang et al., 2014b; Van Leeuwen et al., 2011), permet d'enrichir les mesures faites dans le cadre d'inventaires de gestion ou d'évaluation de la ressource. La phase d'acquisition, notamment sa lourdeur restent des freins importants au développement des utilisations en opérationnel, peut-être plus que le traitement des données 3D. L'enjeu de ce travail est de franchir un nouveau pas dans l'estimation de la qualité des troncs à partir de données 3D, avec des retombées potentielles dans le domaine de l'inventaire ou dans celui de l'estimation de la qualité des bois ronds ou de l'optimisation de la transformation.

Ainsi l'objectif initial de cette thèse était de proposer une suite d'algorithmes innovants permettant de traiter des données 3D décrivant avec précision la surface d'un tronc ou d'un billon afin d'y caractériser les défauts locaux. Ces défauts révélés par des irrégularités de surface de l'écorce sont souvent liés à des défauts internes comme les nœuds des branches englobés par la croissance radiale.

Cet objectif a été atteint. La méthode mise au point se décompose en trois principales étapes : (i) : détection des défauts ; (ii) : identification des types des défauts ; (iii) : caractérisation des défauts.

L'objectif de la première étape était de repérer rapidement les zones sur la surface où une analyse du relief révèle une singularité susceptible de correspondre à un défaut. Dans cette étape, nous avons implémenté une modification de l'algorithme proposé par Kerautret et al. (2016) pour calculer la ligne centrale du tronc ou du billon. La modification comprend des adaptations à la volumétrie des données liées à la taille des troncs et au bruit des données pour créer une ligne centrale lissée.

Ensuite, nous avons développé un algorithme original pour détecter les irrégularités sur la surface du tronc. L'algorithme est basé sur un patch des points voisins pour calculer la distance

de référence à la ligne centrale de chaque point dans le nuage de points. Une fois la distance de référence calculée pour chaque point, nous avons implémenté l'algorithme de Rosin (Rosin, 2001) pour trouver le seuil optimal de la différence entre les distances à la ligne centrale et les distances de référence du nuage de points. Ce seuil permet de détecter les points susceptibles d'appartenir aux défauts de manière assez lâche afin de n'oublier aucun défaut même si en contre-partie des zones d'écorce sont détectés sans être de vrais défauts. Nous avons ensuite utilisé la méthode de clustering par la distance euclidienne implémentée dans la bibliothèque PCL (Rusu et Cousins, 2011) pour regrouper ces points en des zones étant susceptibles d'être des défauts. Dans une expérimentation menée sur 10 billons de 7 espèces, le résultat de l'algorithme a été comparé avec une référence à dire d'expert. L'exactitude du classement a été évaluée par la mesure  $F_1$  construite à partir de la matrice de confusion et dont le maximum est 1. Nous avons obtenu une valeur moyenne de  $F_1$  de 0.76.

La deuxième étape consistait à identifier le type de défaut de chaque zone détectée lors de la première étape. Cette étape a deux objectifs principaux. Le premier objectif est de confirmer la pertinence de la détection et d'enlever les faux positifs (les zones ne correspondant à aucun défaut). Le deuxième objectif est de classer les défauts selon leur origine biologique définissant leur type. Dans cette étape, nous avons choisi des caractéristiques du défaut telles que l'espèce, les invariants du moment de Hu (Hu, 1962), la dimension du défaut, les valeurs propres en lien avec les axes principaux du défaut, et la moyenne et l'écart-type de la différence entre les distances à la ligne centrale et les distances de référence à la ligne centrale. Dans l'expérimentation menée sur 26 troncs de 6 espèces, une méthode d'apprentissage automatique, en l'occurrence les forêts aléatoires de Breiman (2001), i.e. Random Forests a été adoptée, se basant sur ces caractéristiques. Le Random Forests a été entraîné avec 425 exemples répartis en 4 classes de défaut et une classe d'écorce. Ensuite, sur une base de test comprenant 185 défauts, l'algorithme a identifié correctement 163 défauts et la valeur moyenne de la mesure  $F_1$  était de 0.84, en dépit d'une base d'apprentissage qui pourrait être incrémentée.

La troisième étape avait pour but de déduire les caractéristiques des défauts telles que leur position et leurs dimensions. Nous avons développé des méthodes permettant d'estimer automatiquement ces caractéristiques. Ces méthodes sont différentes selon le type identifié dans la deuxième étape. L'expérimentation sur la quantification des défauts a été menée sur 3 troncs de chêne et 3 troncs de hêtre. Le résultat de l'estimation des dimensions de défauts est apparu prometteur avec une médiane des différences entre les dimensions horizontaux estimés et observés étant de 10,5 mm et celle des dimensions verticaux de 6,5 mm. Cependant, les dimensions sont souvent surestimées et la distribution de la différence entre les dimensions estimées et observées est assez dispersée. Cela reflète surtout le fait que la délimitation de la zone du défaut diffère entre un opérateur humain et l'algorithme proposé qui ne se base que sur le relief.

Cette méthode débouche sur une application opérationnelle implémentée dans un plugin de la plateforme Computree. Ce plugin permet de sortir non seulement les caractéristiques de défauts

évoquées mais aussi une visualisation en couleur des défauts selon leur type.

Les perspectives principales pour ce plugin sont l'intégration des étapes de prétraitement (traitement de bruit et maillage) et des améliorations algorithmiques envisageables pour développer un système informatique d'aide à la décision. Ce système permettra d'assister les procédures conventionnelles de classement selon les normes en vigueur ou d'alimenter la chaîne d'information pouvant servir à optimiser la transformation.

# Bibliographie

- AFNOR (1999a) EN 1927-1 - classement qualitatif des bois ronds résineux - partie 1 : épicéas et sapins qualitative classification of softwood round timber - part 1 : Spruces and firs
- AFNOR (1999b) EN 1927-2 - classement qualitatif des bois ronds résineux - partie 2 : pins qualitative classification of softwood round timber - part 2 : Pines
- AFNOR (1999c) EN 1927-3 - classement qualitatif des bois ronds résineux - partie 3 : mélèzes et douglas qualitative classification of softwood round timber - part 3 : Larches and douglas fir
- AFNOR (2012a) EN 1316-1 - bois ronds feuillus - classement qualitatif - partie 1 : Chêne et hêtre hardwood round timber - qualitative classification - part 1 : Oak and beech
- AFNOR (2012b) EN 1316-2 - bois ronds feuillus - classement qualitatif - partie 1 : Peuplier hardwood round timber - qualitative classification - part 2 : Poplar
- Aguilera C, Sanchez R, Baradit E (2008) Detection of knots using x-ray tomographies and deformable contours with simulated annealing. *Wood Res* 53 :57–66
- Azimi-Sadjadi MR, Shaikh MA, Tian B, Eis KE, Reinke D (1996) Neural network-based cloud detection/classification using textural and spectral features. In : *Geoscience and Remote Sensing Symposium, 1996. IGARSS'96. 'Remote Sensing for a Sustainable Future.'*, International, IEEE, vol 2, pp 1105–1107
- Bauer U, Polthier K (2009) Generating parametric models of tubes from laser scans. *Computer-Aided Design* 41(10) :719–729
- Bentley JL (1975) Multidimensional binary search trees used for associative searching. *Communications of the ACM* 18(9) :509–517
- Bienert A, Scheller S, Keane E, Mohan F, Nugent C (2007) Tree detection and diameter estimations by analysis of forest terrestrial laserscanner point clouds. In : *ISPRS workshop on laser scanning*, vol 2007, pp 50–55
- Bienert A, Hess C, Maas H, Von Oheimb G (2014) A voxel-based technique to estimate the volume of trees from terrestrial laser scanner data. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 40(5) :101

- Bishop CM (1995) *Neural networks for pattern recognition*. Oxford university press
- Blanchette D, Fournier RA, Luther JE, Côté JF (2015) Predicting wood fiber attributes using local-scale metrics from terrestrial lidar data : A case study of newfoundland conifer species. *Forest Ecology and Management* 347 :116–129
- Bradski G (2000) *The OpenCV Library*. Dr Dobb's Journal of Software Tools
- Breiman L (2001) Random forests. *Machine learning* 45(1) :5–32
- Breiman L, Friedman J, Stone CJ, Olshen RA (1984) *Classification and regression trees*. CRC press
- Brolly G, Király G (2009) Algorithms for stem mapping by means of terrestrial laser scanning. *Acta Silvatica et Lignaria Hungarica* 5 :119–130
- Bugnon F (1980) Bases fondamentales pour l'étude de la ramification chez les végétaux. principales modalités de ce phénomène. *Bulletin de la Société Botanique de France Actualités Botaniques* 127(2) :5–14
- Calders K, Newnham G, Burt A, Murphy S, Raunonen P, Herold M, Culvenor D, Avitabile V, Disney M, Armston J, et al. (2015) Nondestructive estimates of above-ground biomass using terrestrial laser scanning. *Methods in Ecology and Evolution* 6(2) :198–208
- Carpenter RD, Jones M (1989) *Defects in hardwood timber*. 678, US Department of Agriculture, Forest Service
- Chang SJ (1992) External and internal defect detection to optimize cutting of hardwood logs and lumber. *Transferring technologies for industry (USA)*
- Chehata N, Guo L, Mallet C (2009) Airborne lidar feature selection for urban classification using random forests. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 38(Part 3) :W8
- Cifuentes R, Van der Zande D, Salas C, Tits L, Farifteh J, Coppin P (2017) Modeling 3d canopy structure and transmitted par using terrestrial lidar. *Canadian Journal of Remote Sensing* 43(2) :124–139
- Colin F, Fontaine F, Ningre F (2007) Gourmands et autres épïcormiques du chêne et du hêtre. partie i : un renouvellement des concepts pour une réactivation des recherches. *Forêt Wallonne* 87 :36–48
- Colin F, Mechergui R, Dhôte JF, Fontaine F (2010a) Epicormic ontogeny on quercus petraea trunks and thinning effects quantified with the epicormic composition. *Annals of forest science* 67(8) :813–813
- Colin F, Mothe F, Freyburger C, Morisset JB, Leban JM, Fontaine F (2010b) Tracking rameal traces in sessile oak trunks with X-ray computer tomography : Biological bases, preliminary results and perspectives. *Trees - Structure and Function* 24(5) :953–967, doi: 10.1007/s00468-010-0466-1

- Cortes C, Vapnik V (1995) Support-vector networks. *Machine learning* 20(3) :273–297
- Côté JF, Fournier Ra, Egli R (2011) An architectural model of trees to estimate forest structural attributes using terrestrial LiDAR. *Environmental Modelling and Software* 26(6) :761–777, doi: 10.1016/j.envsoft.2010.12.008, URL <http://dx.doi.org/10.1016/j.envsoft.2010.12.008>
- Danson FM, Hetherington D, Morsdorf F, Koetz B, Allgower B (2006) Three-dimensional forest canopy structure from terrestrial laser scanning. Koukal and Schneider [13] pp 61–65
- Dassot M, Colin A, Santenoise P, Fournier M, Constant T (2012) Terrestrial laser scanning for measuring the solid wood volume, including branches, of adult standing trees in the forest environment. *Computers and Electronics in Agriculture* 89 :86–93
- Devroye L, Györfi L, Lugosi G (2013) *A probabilistic theory of pattern recognition*, vol 31. Springer Science & Business Media
- DGtal (2016) DGtal : Digital Geometry tools and algorithms library. URL <http://dgtal.org>, [Online; accessed 31-August-2018]
- Díaz-Uriarte R, De Andres SA (2006) Gene selection and classification of microarray data using random forest. *BMC bioinformatics* 7(1) :3
- Duchateau E, Longuetaud F, Mothe F, Ung C, Auty D, Achim A (2013) Modelling knot morphology as a function of external tree and branch attributes. *Canadian journal of forest research* 43(3) :266–277
- Fischler MA, Bolles RC (1981) Random sample consensus : a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM* 24(6) :381–395
- Flann (2016) Flann library - fast library for approximate nearest neighbors. URL <http://www.cs.ubc.ca/research/flann>, [Online; accessed 31-August-2018]
- Fonseca MA (2005) *The measurement of roundwood : methodologies and conversion ratios*. CABI
- Fontaine F, Jarret P, Druelle J (2002) Etude et suivi des bourgeons épïcormiques a l’origine des gourmands chez le chêne sessile. *Revue forestière française* 54 :337–356
- García M, Danson FM, Riano D, Chuvieco E, Ramirez FA, Bandugula V (2011) Terrestrial laser scanning to estimate plot-level forest canopy fuel properties. *International journal of applied earth observation and geoinformation* 13(4) :636–645
- Gislason PO, Benediktsson JA, Sveinsson JR (2006) Random forests for land cover classification. *Pattern Recognition Letters* 27(4) :294–300
- Gough B (2009) *GNU scientific library reference manual*. Network Theory Ltd.

- Guan H, Yu Y, Ji Z, Li J, Zhang Q (2015) Deep learning-based tree classification using mobile lidar data. *Remote Sensing Letters* 6(11) :864–873
- Guo L, Chehata N, Mallet C, Boukir S (2011) Relevance of airborne lidar and multispectral image data for urban scene classification using random forests. *ISPRS Journal of Photogrammetry and Remote Sensing* 66(1) :56–66
- Hackenberg J, Wassenberg M, Spiecker H, Sun D (2015) Non destructive method for biomass prediction combining tfs derived tree volume and wood density. *Forests* 6(4) :1274–1300
- Hauglin M, Astrup R, Gobakken T, Næsset E (2013) Estimating single-tree branch biomass of norway spruce with terrestrial laser scanning using voxel-based and crown dimension features. *Scandinavian journal of forest research* 28(5) :456–469
- Heinzel J, Huber MO (2016) Detecting tree stems from volumetric tfs data in forest environments with rich understory. *Remote Sensing* 9(1) :9
- Henning JG, Radtke PJ (2006a) Detailed stem measurements of standing trees from ground-based scanning lidar. *Forest Science* 52(1) :67–80
- Henning JG, Radtke PJ (2006b) Ground-based laser imaging for assessing three-dimensional forest canopy structure. *Photogrammetric Engineering & Remote Sensing* 72(12) :1349–1358
- Hilker T, Frazer GW, Coops NC, Wulder MA, Newnham GJ, Stewart JD, van Leeuwen M, Culvenor DS (2013) Prediction of wood fiber attributes from lidar-derived forest canopy indicators. *Forest Science* 59(2) :231–242
- Holopainen M, Vastaranta M, Kankare V, Rätty M, Vaaja M, Liang X, Yu X, Hyyppä J, Hyyppä H, Viitala R, et al. (2011) Biomass estimation of individual trees using stem and crown diameter tfs measurements. *ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 3812 :91–95
- Hopkinson C, Chasmer L, Young-Pow C, Treitz P (2004) Assessing forest metrics with a ground-based scanning lidar. *Canadian Journal of Forest Research* 34(3) :573–583
- Hu MK (1962) Visual pattern recognition by moment invariants. *IRE transactions on information theory* 8(2) :179–187
- Jourez B, de Wauters P, Bienfait O (2010) Le classement des bois feuillus sur pied. *Silva Belgica* 117 :1–12
- Jovančević I, Pham HH, Orteu JJ, Gilblas R, Harvent J, Maurice X, Brèthes L (2017) 3d point cloud analysis for detection and characterization of defects on airplane exterior surface. *Journal of Nondestructive Evaluation* 36(4) :74
- Kakde HM (2005) Range searching using kd tree. Florida State University

- Kankare V, Holopainen M, Vastaranta M, Puttonen E, Yu X, Hyyppä J, Vaaja M, Hyyppä H, Alho P (2013) Individual tree biomass estimation using terrestrial laser scanning. *ISPRS Journal of Photogrammetry and Remote Sensing* 75 :64–75
- Kankare V, Joensuu M, Vauhkonen J, Holopainen M, Tanhuanpää T, Vastaranta M, Hyyppä J, Hyyppä H, Alho P, Rikala J, et al. (2014) Estimation of the timber quality of scots pine with terrestrial laser scanning. *Forests* 5(8) :1879–1895
- Kelbe D, Romanczyk P, van Aardt J, Cawse-Nicholson K (2013) Reconstruction of 3d tree stem models from low-cost terrestrial laser scanner data. In : *Laser Radar Technology and Applications XVIII, International Society for Optics and Photonics*, vol 8731, p 873106
- Kelbe D, van Aardt J, Romanczyk P, van Leeuwen M, Cawse-Nicholson K (2015) Single-scan stem reconstruction using low-resolution terrestrial laser scanner data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 8(7) :3414–3427
- Kerautret B, Krähenbühl A, Debled-Rennesson I, Lachaud JO (2015) 3d geometric analysis of tubular objects based on surface normal accumulation. In : *Image Analysis and Processing—ICIAP 2015, Springer*, pp 319–331
- Kerautret B, Krähenbühl A, Debled-Rennesson I, Lachaud JO (2016) Centerline detection on partial mesh scans by confidence vote in accumulation map. In : *Pattern Recognition (ICPR), 2016 23rd International Conference on, IEEE*, pp 1376–1381
- Kirbas C, Quek F (2004) A review of vessel extraction techniques and algorithms. *CSUR* 36(2) :81–121
- Klemmt HJ, Seifert T, Seifert S, Kunneke A, Wessels B, Pretzsch H (2010) Assessment of branchiness in a *Pinus pinaster* plantation by terrestrial laser scanner data as a link between exterior and interior wood properties. In : *Proceedings of the 10th International Conference on LiDAR Applications for Assessing Forest Ecosystems, University of Freiburg, Germany*
- Krähenbühl A, Kerautret B, Debled-Rennesson I, Longuetaud F, Mothe F (2012) Knot detection in x-ray ct images of wood. In : *International Symposium on Visual Computing, Springer*, pp 209–218
- Krähenbühl A, Roussel JR, Kerautret B, Debled-Rennesson I, Mothe F, Longuetaud F (2016) Robust knot segmentation by knot pith tracking in 3d tangential images. In : *International Conference on Computer Vision and Graphics, Springer*, pp 581–593
- Kretschmer U, Kirchner N, Morhart C, Spiecker H (2013) A new approach to assessing tree stem quality characteristics using terrestrial laser scans. *Silva Fennica* 47(5) :1–14, doi: 10.14214/sf.1071
- Krähenbühl A, Kerautret B, Debled-Rennesson I, Mothe F, Longuetaud F (2014) Knot Segmentation in 3d CT Images of Wet Wood. *Pattern Recognition*



- Kwak DA, Cui G, Lee WK, Cho HK, Jeon SW, Lee SH (2014) Estimating plot volume using lidar height and intensity distributional parameters. *International journal of remote sensing* 35(13) :4601–4629
- Laefer DF, Truong-Hong L, Carr H, Singh M (2014) Crack detection limits in unit based masonry with terrestrial laser scanning. *NDT & E International* 62 :66–76
- LeCun Y, Bengio Y, Hinton G (2015) Deep learning. *nature* 521(7553) :436
- Lefsky MA, McHale MR (2008) Volume estimates of trees with complex architecture from terrestrial laser scanning. *Journal of Applied Remote Sensing* 2(1) :023521
- Lesage D, Angelini ED, Bloch I, Funke-Lea G (2009) A review of 3d vessel lumen segmentation techniques : Models, features and extraction schemes. *Medical Image Analysis* 13(6) :819–845, doi: 10.1016/j.media.2009.07.011
- Li PLP, aL Abbott, Schmoldt D (1996) Automated analysis of CT images for the inspection of hardwood logs. In : *Proceedings of International Conference on Neural Networks (ICNN'96)*, IEEE, vol 3, pp 1744–1749, doi: 10.1109/ICNN.1996.549164
- Liang X, Litkey P, Hyypä J, Kukko A, Kaartinen H, Holopainen M (2008) Plot-level trunk detection and reconstruction using one-scan-mode terrestrial laser scanning data. In : *Earth Observation and Remote Sensing Applications, 2008. EORSA 2008. International Workshop on*, IEEE, pp 1–5
- Liang X, Hyypä J, Kaartinen H, Holopainen M, Melkas T (2012) Detecting changes in forest structure over time with bi-temporal terrestrial laser scanning data. *ISPRS International Journal of Geo-Information* 1(3) :242–255
- Liang X, Hyypä J, Kukko A, Kaartinen H, Jaakkola A, Yu X (2014a) The use of a mobile laser scanning system for mapping large forest plots. *IEEE Geoscience and Remote Sensing Letters* 11(9) :1504–1508
- Liang X, Kankare V, Yu X, Hyypä J, Holopainen M (2014b) Automated stem curve measurement using terrestrial laser scanning. *Geoscience and Remote Sensing, IEEE Transactions on* 52(3) :1739–1748
- Litkey P, Liang X, Kaartinen H, Hyypä J, Kukko A, Holopainen M (2008) Single-scan tls methods for forest parameter retrieval. *Proceedings of SilviLaser 2008* :8th
- Longuetaud F, Mothe F, Kerautret B, Krähenbü A, Hory L, Leban JM, Debled-Rennesson I (2012) Automatic knot detection and measurements from X-ray CT images of wood : A review and validation of an improved algorithm on softwood samples. *COMPAG* 85 :77–89
- López-García F, Andreu-García G, Blasco J, Aleixos N, Valiente JM (2010) Automatic detection of skin defects in citrus fruits using a multivariate image analysis approach. *Computers and Electronics in Agriculture* 71(2) :189–197

- Lowe DG (2004) Distinctive image features from scale-invariant keypoints. *International journal of computer vision* 60(2) :91–110
- Luther JE, Skinner R, Fournier RA, van Lier OR, Bowers WW, Coté JF, Hopkinson C, Moulton T (2013) Predicting wood quantity and quality attributes of balsam fir and black spruce using airborne laser scanner data. *Forestry* 87(2) :313–326
- Maas HG, Bienert A, Scheller S, Keane E (2008) Automatic forest inventory parameter determination from terrestrial laser scanner data. *International journal of remote sensing* 29(5) :1579–1593
- Maltamo M, Peuhkurinen J, Malinen J, Vauhkonen J, Packalén P, Tokola T, et al. (2009) Predicting tree attributes and quality characteristics of scots pine using airborne laser scanning data. *Silva Fennica* 43(3) :507–521
- Maturana D, Scherer S (2015) 3d convolutional neural networks for landing zone detection from lidar. In : *Robotics and Automation (ICRA), 2015 IEEE International Conference on*, IEEE, pp 3471–3478
- McCullough DG, Heyd RL, O'Brien JG, Marquette M (2005) Biology and management of beech bark disease. *Michigan State University Extension Bulletin* E-2746
- Meadows JS, Burkhardt E (2001) Epicormic branches affect lumber grade and value in willow oak. *Southern Journal of Applied Forestry* 25(3) :136–141
- Mengesha T, Hawkins M, Nieuwenhuis M (2015) Validation of terrestrial laser scanning data using conventional forest inventory methods. *European journal of forest research* 134(2) :211–222
- Moorthy I, Miller JR, Hu B, Chen J, Li Q (2008) Retrieving crown leaf area index from an individual tree using ground-based lidar data. *Canadian Journal of Remote Sensing* 34(3) :320–332
- Morel J, Bac A, Véga C (2018) Surface reconstruction of incomplete datasets : A novel poisson surface approach based on csrbf. *Computers & Graphics* 74 :44–55
- Moskal LM, Zheng G (2011) Retrieving forest inventory variables with terrestrial laser scanning (tls) in urban heterogeneous forest. *Remote Sensing* 4(1) :1–20
- Muja M, Lowe DG (2014) Scalable Nearest Neighbor Algorithms for High Dimensional Data. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 36
- Mutanga O, Adam E, Cho MA (2012) High density biomass estimation for wetland vegetation using worldview-2 imagery and random forest regression algorithm. *International Journal of Applied Earth Observation and Geoinformation* 18 :399–406
- Nguyen VT, Kerautret B, Debled-Rennesson I, Colin F, Piboule A, Constant T (2016a) Algorithms and implementation for segmenting tree log surface defects. In : *International Workshop on Reproducible Research in Pattern Recognition*, Springer, pp 150–166

- Nguyen VT, Kerautret B, Debled-Rennesson I, Colin F, Piboule A, Constant T (2016b) Online demonstration of the segmentation of defects on log surface from terrestrial lidar data. URL [http://ipol-geometry.loria.fr/~kerautre/ipol\\_demo/TrunkDefaultMeasure](http://ipol-geometry.loria.fr/~kerautre/ipol_demo/TrunkDefaultMeasure), [Online; accessed 31-August-2018]
- Nguyen VT, Kerautret B, Debled-Rennesson I, Colin F, Piboule A, Constant T (2016c) Segmentation of defects on log surface from terrestrial lidar data. In : Pattern Recognition (ICPR), 2016 23rd International Conference on, IEEE, pp 3168–3173
- Nguyen VT, Constant T, Colin F (2018a) An innovative and automated method for characterizing wood defects on trunk surface using high density 3d terrestrial lidar data, (soumis)
- Nguyen VT, Constant T, Kerautret B, Debled-Rennesson I, Colin F (2018b) A machine-learning approach for classifying defects on tree trunk using terrestrial LiDAR, (à soumettre)
- Nguyen VT, Kerautret B, Debled-Rennesson I, Colin F, Piboule A, Constant T (2018c) Branches and defects segmentation on the surface of tree trunk from partial or full terrestrial lidar data, (en projet)
- Niemeyer J, Rottensteiner F, Soergel U (2014) Contextual classification of lidar data and building object detection in urban areas. *ISPRS journal of photogrammetry and remote sensing* 87 :152–165
- Olofsson K, Holmgren J, Olsson H (2014) Tree stem and height measurements using terrestrial laser scanning and the ransac algorithm. *Remote sensing* 6(5) :4323–4344
- Osborne NL, Maguire DA (2015) Modeling knot geometry from branch angles in douglas-fir (*pseudotsuga menziesii*). *Canadian Journal of Forest Research* 46(2) :215–224
- Othmani A, Voon LFLY, Stolz C, Piboule A (2013) Single tree species classification from terrestrial laser scanning data for forest inventory. *Pattern Recognition Letters* 34(16) :2144–2150
- Pfeifer N, Gorte B, Winterhalder D (2004) Automatic reconstruction of single trees from terrestrial laser scanner data. In : Proceedings of 20th ISPRS Congress, pp 114–119
- Pokharel B, Groot A, Pitt DG, Woods M, Dech JP (2016) Predictive modeling of black spruce (*picea mariana* (mill.) bsp) wood density using stand structure variables derived from airborne lidar data in boreal forests of ontario. *Forests* 7(12) :311
- Popescu SC, Wynne RH, Nelson RF (2002) Estimating plot-level tree heights with lidar : local filtering with a canopy-height based variable window size. *Computers and Electronics in Agriculture* 37(1) :71–95
- Provost F, Kohavi R (1998) Glossary of terms. *Journal of Machine Learning* 30(2-3) :271–274
- Puissant A, Rougier S, Stumpf A (2014) Object-oriented mapping of urban trees using random forest classifiers. *International Journal of Applied Earth Observation and Geoinformation* 26 :235–245

- Pyörälä J, Kankare V, Vastaranta M, Rikala J, Holopainen M, Sipi M, Hyyppä J, Uusitalo J (2017) Comparison of terrestrial laser scanning and x-ray scanning in measuring scots pine (*pinus sylvestris* l.) branch structure. *Scandinavian Journal of Forest Research* pp 1–8
- Racko V (2013) Verify the accuracy of estimation the model between dimensional characteristics of branch scar and the location of the knot in the beech trunk. *Annals of Warsaw Univ of Life Sciences-SGGW Forestry & Wood Technology* 84
- Raumonen P, Kaasalainen M, Åkerblom M, Kaasalainen S, Kaartinen H, Vastaranta M, Holopainen M, Disney M, Lewis P (2013) Fast automatic precision tree models from terrestrial laser scanner data. *Remote Sensing* 5(2) :491–520
- Rodriguez-Galiano VF, Ghimire B, Rogan J, Chica-Olmo M, Rigol-Sanchez JP (2012) An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing* 67 :93–104
- Rosin PL (2001) Unimodal thresholding. *Pattern recognition* 34(11) :2083–2096
- Rusu RB, Cousins S (2011) 3D is here : Point Cloud Library (PCL). In : *IEEE International Conference on Robotics and Automation (ICRA)*, Shanghai, China
- Schafer ME (2000) Ultrasound for defect detection and grading in wood and lumber. In : *Ultrasonics Symposium, 2000 IEEE, IEEE, vol 1*, pp 771–778
- Schütt C, Aschoff T, Winterhalder D, Thies M, Kretschmer U, Spiecker H (2004) Approaches for recognition of wood quality of standing trees based on terrestrial laserscanner data. *Laser-scanners for forest and landscape assessment Proceedings of the ISPRS working group VIII/2, Freiburg, Germany International Archives of Photogrammetry, Remote Sensing, and Spatial Information Sciences* 36 :179–182
- Seidel D, Ammer C, Puettmann K (2015) Describing forest canopy gaps efficiently, accurately, and objectively : new prospects through the use of terrestrial laser scanning. *Agricultural and forest meteorology* 213 :23–32
- Simonse M, Aschoff T, Spiecker H, Thies M (2003) Automatic determination of forest inventory parameters using terrestrial laserscanning. In : *Proceedings of the ScandLaser Scientific Workshop on Airborne Laser Scanning of Forests, vol 2003*, pp 252–258
- Sinha SK, Fieguth PW (2006) Neuro-fuzzy network for the classification of buried pipe defects. *Automation in Construction* 15(1) :73–83
- Soudarissanane S, Lindenbergh R, Menenti M, Teunissen P (2009) Incidence angle influence on the quality of terrestrial laser scanning points. In : *Proceedings ISPRS Workshop Laserscanning 2009, 1-2 Sept 2009, Paris, France, ISPRS*

- Stängle SM, Brüchert F, Kretschmer U, Spiecker H, Sauter UH (2013) Clear wood content in standing trees predicted from branch scar measurements with terrestrial lidar and verified with x-ray computed tomography 1. *Canadian Journal of Forest Research* 44(2) :145–153
- Suvdaa B, Ahn J, Ko J (2012) Steel surface defects detection and classification using SIFT and voting strategy. *International Journal of Software Engineering and its Applications* 6(2) :161–166
- Tague-Sutcliffe J (1992) The pragmatics of information retrieval experimentation, revisited. *Information Processing & Management* 28(4) :467 – 490, doi: [https://doi.org/10.1016/0306-4573\(92\)90005-K](https://doi.org/10.1016/0306-4573(92)90005-K), URL <http://www.sciencedirect.com/science/article/pii/030645739290005K>, special Issue : Evaluation Issues in Information Retrieval
- Tankyevych O, Talbot H, Passat N, Musacchio M, Lagneau M (2011) Angiographic image analysis. In : *Medical Image Processing*, Springer, pp 115–144
- Terzopoulos D, Fleischer K (1988) Deformable models. *The visual computer* 4(6) :306–331
- Thies M, Spiecker H (2004) Evaluation and future prospects of terrestrial laser scanning for standardized forest inventories. *Forest* 2(2.2) :1
- Thomas L, Mili L (2007) A robust gm-estimator for the automated detection of external defects on barked hardwood logs and stems. *Signal Processing, IEEE Transactions on* 55(7) :3568–3576
- Thomas L, Shaffer Ca, Mili L, Thomas E (2007) Automated detection of severe surface defects on barked hardwood logs. *Forest Products Journal* 57(April) :50–56
- Thomas L, Thomas E, et al. (2010) A graphical automated detection system to locate hardwood log surface defects using high-resolution three-dimensional laser scan data. In : *Proceedings of the 17th Central Hardwood Forest Conference*, pp 5–7
- Thomas RE (2009) Modeling the relationships among internal defect features and external Appalachian hardwood log defect indicators. *Silva Fennica* 43(3) :447–456
- Thomas RE, Thomas L (2013) Using parallel computing methods to improve log surface defect detection methods. In : In : Ross, Robert J.; Wang, Xiping, eds. *Proceedings, 18th International Nondestructive Testing and Evaluation of Wood Symposium; 2013 September 24-27; Madison, WI. Gen. Tech. Rep. FPL-226. Madison, WI: US Department of Agriculture, Forest Service, Forest Products Laboratory : 196-205.*, pp 196–205
- Torkaman J, Vaziri M, Sandberg D, Mohammadi Limaie S (2018) Relationship between branch-scar parameters and knot features of oriental beech (*fagus orientalis libsky*). *Wood Material Science & Engineering* 13(2) :117–120
- Van Leeuwen M, Hilker T, Coops NC, Frazer G, Wulder MA, Newnham GJ, Culvenor DS (2011) Assessment of standing wood and fiber quality using ground and airborne laser scanning : a review. *Forest Ecology and Management* 261(9) :1467–1478

- Vastaranta M, et al. (2012) Forest mapping and monitoring using active 3d remote sensing. *Dissertationes Forestales*
- Voegtle T, Wakaluk S (2009) Effects on the measurements of the terrestrial laser scanner hds 6000 (leica) caused by different object materials. *Proceedings of ISPRS Work 38(2009)* :68–74
- Wang D, Hollaus M, Puttonen E, Pfeifer N (2016) Fast and robust stem reconstruction in complex environments using terrestrial laser scanning. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences* 41
- Watt P, Donoghue D (2005) Measuring forest structure with terrestrial laser scanning. *International Journal of Remote Sensing* 26(7) :1437–1446
- Wernsdörfer H, Constant T, Mothe F, Badia M, Nepveu G, Seeling U (2005) Detailed analysis of the geometric relationship between external traits and the shape of red heartwood in beech trees (*fagus sylvatica* l.). *Trees* 19(4) :482–491
- Wold S, Esbensen K, Geladi P (1987) Principal component analysis. *Chemometrics and intelligent laboratory systems* 2(1-3) :37–52
- Yao T, Yang X, Zhao F, Wang Z, Zhang Q, Jupp D, Lovell J, Culvenor D, Newnham G, Ni-Meister W, et al. (2011) Measuring forest structure and biomass in new england forest stands using echidna ground-based lidar. *Remote sensing of Environment* 115(11) :2965–2974
- You L, Tang S, Song X, Lei Y, Zang H, Lou M, Zhuang C (2016) Precise measurement of stem diameter by simulating the path of diameter tape from terrestrial laser scanning data. *Remote Sensing* 8(9) :717
- Yu X, Liang X, Hyypä J, Kankare V, Vastaranta M, Holopainen M (2013) Stem biomass estimation based on stem reconstruction from terrestrial laser scanning point clouds. *Remote sensing letters* 4(4) :344–353
- Yuan F, Sawaya KE, Loeffelholz BC, Bauer ME (2005) Land cover classification and change analysis of the twin cities (minnesota) metropolitan area by multitemporal landsat remote sensing. *Remote sensing of Environment* 98(2) :317–328
- Zande DVd, Jonckheere I, Stuckens J, Verstraeten WW, Coppin P (2008) Sampling design of ground-based lidar measurements of forest canopy structure and its effect on shadowing. *Canadian Journal of Remote Sensing* 34(6) :526–538
- Zhu D, Connors RW, Schmoldt DL, Araman Pa (1996) A prototype vision system for analyzing CT imagery of hardwood logs. *IEEE Transactions on Systems, Man, and Cybernetics, Part B : Cybernetics* 26(4) :522–532, doi: 10.1109/3477.517028
- Zhu XX, Tuia D, Mou L, Xia GS, Zhang L, Xu F, Fraundorfer F (2017) Deep learning in remote sensing : a comprehensive review and list of resources. *IEEE Geoscience and Remote Sensing Magazine* 5(4) :8–36



**Title :** Assessment of roundwood and tree quality from terrestrial Lidar data

**Keywords :** forest inventory, 3D scanning, external defects

**Abstract :**

*The Terrestrial Laser Scanning (TLS) technology is booming in forestry applications. It generates a cloud of tens of millions of 3D points corresponding to the obstacles encountered by a laser beam scanning a scene from several points of view. Its advantage lies in its ability to describe a forest plot or an individual tree by a point cloud with a high level of detail unusual for such a large object. This level of detail allows a fine analysis of the trunk surface to detect the defects that play an essential role in wood quality. In fact, several studies have shown that 80% of internal defects correspond to traces on the trunk surface, for example, branch scars.*

*The objective of the thesis is to develop a method to automatically quantify defects on trunks of standing trees or logs, using TLS data. The defects included branches, branch scars, burls, bud cluster, and picots.*

*The development is completed through 3 steps : (1) defect detection, (2) defect identification and (3) defect characterization.*

*The detection of defects is based on an analysis of the bark roughness using an estimation of the reference distance to the trunk centerline. The principle of centerline computation is to find the points where the surface normal vectors converge. The reference distance of each point from the surface to this centerline is computed from the distribution of its neighboring points belonging to the elongated patch along the main axis of the trunk. The detection of defects is based on the idea that, for one point belonging to a defect, the difference between the real distance and the reference distance to the centerline exceeds a certain threshold value. The reference distance computation takes also into account the local roughness of the bark. The points supposed to correspond to a defect are obtained by an automatic thresholding of the statistical distribution of the differences. The defect detection method can take into account a large variety of shapes of the trunk or log ; it uses a small number of parameters not sensitive to the nature of the bark.*

*Then, the identification of defects is based on a machine learning approach using the Random Forests classifier, which not only classifies defects by type but also reduces the number of false positives resulting from the first step.*

*The characterization of defects includes the estimation of their position and dimensions associated with their type.*

*Experiments on logs or trunks of different species such as oak, beech, wild cherry, pine, fir, and spruce showed that we can detect defects with a size as small as 5 mm with sufficiently dense TLS data. When compared at the surface element level, the results were well consistent with the data observed visually on the meshes created from the TLS point cloud. The criterion chosen to evaluate the classification performance is the  $F_1$  score, which is the harmonic mean of the precision and the recall, two variables calculated from the false and true positives. Concerning the detection step,  $F_1$  was 0.76. The second step correctly classified 163 defects*



out of 185 but induced 75 false identifications of mostly small defects. The  $F_1$  score was 0.84.

The latest experiments on oaks and beeches, designed to automatically evaluate the quality of standing trees, showed that the proposed method was robust and promising for these two species with very different bark structures.

Compared to previous work on the use of TLS data for similar objectives, the developed tool considerably improves the automatic quantification of external defects and estimation of standing tree quality. The main perspectives of this work are : improve the variability of the defects contained in the training set and possibly separate the training set by species, and extend this tool to take into account other types of external defects, such as cracks, and the global attributes of the trunk or log to estimate log quality on the wood yard or in the forest.

**Titre :** Estimation de la qualité de bois ronds et d'arbres sur pied par Lidar terrestre

**Mots clés :** inventaire forestier, scans 3D, défauts externes

**Résumé :**

*La technologie LiDAR (Light Detection And Ranging) terrestre (TLS) est une méthode de télédétection en plein essor dans le domaine forestier. Elle génère un nuage de dizaines de millions de points 3D correspondant aux obstacles rencontrés par un faisceau laser qui balaie une scène depuis plusieurs points de vue. Son intérêt réside dans son aptitude à décrire une placette forestière ou un arbre individuel avec un niveau de détail inégalé. Ce niveau de détail permet une analyse fine de la surface du tronc pour en détecter les défauts qui jouent un rôle essentiel dans la qualité du bois puisque des études ont montré que 80% des défauts internes correspondent à des traces à la surface du tronc, par exemple les cicatrices de branches.*

*L'objectif de la thèse est de développer une méthode pour quantifier les défauts sur le tronc d'arbre sur pied ou de bois ronds, à partir de données issues de scans TLS. Les défauts visés sont les branches, les cicatrices de branches, les broussins, les amas de bourgeons et les picots.*

*Le développement de cette méthode a été mené selon 3 étapes : (1) détection des défauts, (2) identification de leur type et (3) caractérisation.*

*La détection des défauts est basée sur une analyse de la rugosité de la surface de l'écorce via une estimation de la distance de référence à la ligne centrale du tronc. La distance de référence de tout point de la surface à cette ligne centrale est calculée à partir de la distribution de ses points voisins appartenant au même patch (bande) allongé selon l'axe principal du tronc. Les points supposés correspondre à des défauts sont obtenus par un seuillage automatique sur la distribution statistique des écarts entre la distance à la ligne centrale et la distance de référence.*

*L'identification des défauts est basée sur une approche d'apprentissage automatique en utilisant le classifieur Random Forests qui non seulement classe les défauts selon leur type mais permet également de réduire le nombre de faux positifs à l'issue de la première étape.*

*La caractérisation des défauts comprend la détermination de leurs position et dimensions qui dépendent de leur type.*

*Des expérimentations réalisées sur des billons ou des troncs de différentes espèces telles que chêne, hêtre, merisier, pin, sapin et épicéa, ont montré que nous pouvons détecter des défauts avec une dimension pouvant descendre jusqu'à 5 mm sur des données TLS suffisamment denses. A l'échelle d'un élément de surface, les résultats sont bien cohérents avec ceux de données observées visuellement sur les maillages créés à partir des nuages de points TLS. Le critère choisi pour évaluer la performance de la classification est la mesure  $F_1$  qui*

*est la moyenne harmonique de la précision et du rappel, deux variables calculées à partir des nombres de faux et vrais positifs. Concernant la détection de défauts,  $F_1$  atteint 0,76.*

*La deuxième étape a correctement classé 163 défauts sur 185 mais a induit 75 fausses identifications de petits défauts majoritairement. La valeur de  $F_1$  est de 0,84.*

*Les dernières expérimentations sur des chênes et des hêtres, destinées à évaluer automatiquement l'ensemble de la méthode sur des arbres sur pied, à partir du nombre et des dimensions des défauts ont montré qu'elle était robuste et prometteuse pour ces deux espèces ayant des structures d'écorce très différentes.*

*Par rapport aux travaux antérieurs sur l'utilisation des données TLS pour des objectifs similaires, l'outil développé permet d'améliorer considérablement la quantification automatique des défauts de surface et donc d'estimer la qualité des grumes sur pied. Les principales perspectives sont : améliorer l'échantillonnage de la variabilité des défauts contenu dans la base d'apprentissage et éventuellement leur séparation par espèce et prendre en compte d'autres types de défauts externes, comme les fentes et les attributs globaux du tronc ou du billon afin d'estimer la qualité de grume en bord de route ou sur parc à grumes.*