Department of Electrical Engineering and Automation

Perception Systems for Autonomous Forest Machinery

Heikki Hyyti

Aalto University

DOCTORAL THESES

Aalto University publication series
DOCTORAL THESES 157/2023

Perception Systems for Autonomous Forest Machinery

Heikki Hyyti

A doctoral thesis completed for the degree of Doctor of Science (Technology) to be defended, with the permission of the Aalto University School of Electrical Engineering, at a public examination held at the lecture hall AS1 of the school (Maarintie 8, Espoo) on 27 October 2023 at 12 noon.

Aalto University
School of Electrical Engineering
Department of Electrical Engineering and Automation
Autonomous Systems

Supervising professor
Prof. Arto Visala, Aalto University, Finland

Thesis advisor
Prof. Ville Lehtola, University of Twente, Netherlands

Preliminary examiners
Prof. Reza Ghabcheloo, Tampere University, Finland
Prof. Thierry Peynot, Queensland University of Technology, Australia

Opponents
Prof. Ola Ringdahl, Umeå University, Sweden
Prof. Thierry Peynot, Queensland University of Technology, Australia

Aalto University publication series
DOCTORAL THESES 157/2023

ISBN 978-952-64-1451-5 (printed)
ISBN 978-952-64-1452-2 (pdf)
ISSN 1799-4934 (printed)
ISSN 1799-4942 (pdf)
http://urn.fi/URN:ISBN:978-952-64-1452-2

Images: The front cover image by Markus Holopainen. All other images by the author.

Unigrafia Oy
Helsinki 2023

Finland

Publication orders (printed book):
heikki@hyyti.fi

Printed matter
4041-0619

Aalto University

Aalto University, P.O. Box 11000, FI-00076 Aalto www.aalto.fi

Abstract

Author
Heikki Hyyti

Name of the doctoral thesis
Perception Systems for Autonomous Forest Machinery

Publisher School of Electrical Engineering

Unit Department of Electrical Engineering and Automation

Series Aalto University publication series DOCTORAL THESES 157/2023

Field of research Automation Technology

Manuscript submitted 6 October 2022 **Date of the defence** 27 October 2023

Permission for public defence granted (date) 25 August 2023 **Language** English

☐ **Monograph** ☒ **Article thesis** ☐ **Essay thesis**

Abstract

A prerequisite for increasing the autonomy of forest machinery is to provide robots with digital situational awareness, including a representation of the surrounding environment and the robot's own state in it. Therefore, this article-based dissertation proposes perception systems for autonomous or semi-autonomous forest machinery as a summary of seven publications. The work consists of several perception methods using machine vision, lidar, inertial sensors, and positioning sensors. The sensors are used together by means of probabilistic sensor fusion. Semi-autonomy is interpreted as a useful intermediary step, situated between current mechanized solutions and full autonomy, to assist the operator.

In this work, the perception of the robot's self is achieved through estimation of its orientation and position in the world, the posture of its crane, and the pose of the attached tool. The view around the forest machine is produced with a rotating lidar, which provides approximately equal-density 3D measurements in all directions. Furthermore, a machine vision camera is used for detecting young trees among other vegetation, and sensor fusion of an actuated lidar and machine vision camera is utilized for detection and classification of tree species. In addition, in an operator-controlled semi-autonomous system, the operator requires a functional view of the data around the robot. To achieve this, the thesis proposes the use of an augmented reality interface, which requires measuring the pose of the operator's head-mounted display in the forest machine cabin. Here, this work adopts a sensor fusion solution for a head-mounted camera and inertial sensors.

In order to increase the level of automation and productivity of forest machines, the work focuses on scientifically novel solutions that are also adaptable for industrial use in forest machinery. Therefore, all the proposed perception methods seek to address a real existing problem within current forest machinery. All the proposed solutions are implemented in a prototype forest machine and field tested in a forest. The proposed methods include posture measurement of a forestry crane, positioning of a freely hanging forestry crane attachment, attitude estimation of an all-terrain vehicle, positioning a head mounted camera in a forest machine cabin, detection of young trees for point cleaning, classification of tree species, and measurement of surrounding tree stems and the ground surface underneath.

Keywords sensors, sensor fusion, Kalman filter, particle filter, lidar, machine vision, inertial measurements, forestry

ISBN (printed) 978-952-64-1451-5 **ISBN (pdf)** 978-952-64-1452-2

ISSN (printed) 1799-4934 **ISSN (pdf)** 1799-4942

Location of publisher Helsinki **Location of printing** Helsinki **Year** 2023

Pages 274 **urn** http://urn.fi/URN:ISBN:978-952-64-1452-2

Aalto-yliopisto

Aalto-yliopisto, PL 11000, 00076 Aalto www.aalto.fi

Tiivistelmä

Tekijä
Heikki Hyyti

Väitöskirjan nimi
Autonomisten metsäkoneiden koneaistijärjestelmät

Julkaisija Sähkötekniikan korkeakoulu

Yksikkö Sähkötekniikan ja automaation laitos

Sarja Aalto University publication series DOCTORAL THESES 157/2023

Tutkimusala Automaatiotekniikka

Käsikirjoituksen pvm 06.10.2022 **Väitöspäivä** 27.10.2023

Väittelyluvan myöntämispäivä 25.08.2023 **Kieli** Englanti

☐ **Monografia** ☒ **Artikkeliväitöskirja** ☐ **Esseeväitöskirja**

Tiivistelmä

Metsäkoneiden autonomia-asteen kasvattaminen edellyttää, että robotilla on digitaalinen tilannetieto sekä ympäristöstä että robotin omasta toiminnasta. Tämän saavuttamiseksi työssä on kehitetty autonomisen tai puoliautonomisen metsäkoneen koneaistijärjestelmiä, jotka hyödyntävät konenäkö-, laserkeilaus- ja inertia-antureita sekä paikannusantureita. Työ liittää yhteen seitsemässä artikkelissa toteutetut havainnointimenetelmät, joissa useiden anturien mittauksia yhdistetään sensorifuusiomenetelmillä. Työssä puoliautonomialla tarkoitetaan hyödyllisiä kuljettajaa avustavia välivaiheita nykyisten mekanisoitujen ratkaisujen ja täyden autonomian välillä.

Työssä esitettävissä autonomisen metsäkoneen koneaistijärjestelmissä koneen omaa toimintaa havainnoidaan estimoimalla koneen asentoa ja sijaintia, nosturin asentoa sekä siihen liitetyn työkalun asentoa suhteessa ympäristöön. Yleisnäkymä metsäkoneen ympärille toteutetaan pyörivällä laserkeilaimella, joka tuottaa lähes vakiotiheyksisiä 3D-mittauksia jokasuuntaisesti koneen ympäristöstä. Nuoret puut tunnistetaan muun kasvillisuuden joukosta käyttäen konenäkökameraa. Lisäksi puiden tunnistamisessa ja puulajien luokittelussa käytetään konenäkökameraa ja laserkeilainta yhdessä sensorifuusioratkaisun avulla. Lisäksi kuljettajan ohjaamassa puoliautonomisessa järjestelmässä kuljettaja tarvitsee toimivan tavan ymmärtää koneen tuottaman mallin ympäristöstä. Työssä tämä ehdotetaan toteutettavaksi lisätyn todellisuuden käyttöliittymän avulla, joka edellyttää metsäkoneen ohjaamossa istuvan kuljettajan lisätyn todellisuuden lasien paikan ja asennon mittaamista. Työssä se toteutetaan kypärään asennetun kameran ja inertia-anturien sensorifuusiona.

Jotta metsäkoneiden automatisaatiotasoa ja tuottavuutta voidaan lisätä, työssä keskitytään uusiin tieteellisiin ratkaisuihin, jotka soveltuvat teolliseen käyttöön metsäkoneissa. Kaikki esitetyt koneaistijärjestelmät pyrkivät vastaamaan todelliseen olemassa olevaan tarpeeseen nykyisten metsäkoneiden käytössä. Siksi kaikki menetelmät on implementoitu prototyyppimetsäkoneisiin ja tulokset on testattu metsäympäristössä. Työssä esitetyt menetelmät mahdollistavat metsäkoneen nosturin, vapaasti riippuvan työkalun ja ajoneuvon asennon estimoinnin, lisätyn todellisuuden lasien asennon mittaamisen metsäkoneen ohjaamossa, nuorten puiden havaitsemisen reikäperkauksessa, ympäröivien puiden puulajien tunnistuksen, sekä puun runkojen ja maanpinnan mittauksen.

Avainsanat anturit, sensorifuusio, Kalman-suodin, partikkelisuodin, laserkeilaus, konenäkö, inertiamittaukset, metsätalous

ISBN (painettu) 978-952-64-1451-5 **ISBN (pdf)** 978-952-64-1452-2

ISSN (painettu) 1799-4934 **ISSN (pdf)** 1799-4942

Julkaisupaikka Helsinki **Painopaikka** Helsinki **Vuosi** 2023

Sivumäärä 274 **urn** http://urn.fi/URN:ISBN:978-952-64-1452-2

Preface

It is my pleasure to invite readers to study the results of the longest, largest, and by far the most challenging project I have undertaken so far. The work for this doctoral thesis was performed during a decade in the Autonomous Systems Group at the Department of Automation and Systems Technology, later merged into the Department of Electrical Engineering and Automation at Aalto University, School of Electrical Engineering, in Espoo, Finland. This journey to the world of science and engineering has been long but worthwhile. I have learned much about sensors, algorithms, robotics, machine perception as well as forest management, heavy machinery, and instrumentation.

The focus of the work is interdisciplinary, and it connects many seemingly separate scientific fields. I hope it will be valuable to other researchers working with probabilistic machine perception or forest robotics. This book is related to instrumentation, sensor modeling, and sensor fusion, as well as navigation, remote sensing, and estimation. The work is also closely related to machine perception, artificial intelligence, and cognitive science, but also knowledge of forest sciences, forest management, and forest machinery is required to understand the framework. Since the focus is wide and probably only a few readers possess sufficient background knowledge from all the related fields, this thesis provides some insight into these disciples so that researchers working only partly on the same topic could also learn from this work.

I could have not completed this thesis without all the help and support I have received from many other people and organizations. Firstly, I wish to thank Professor Arto Visala for all his guidance and supervision. Secondly, I thank all my co-authors in the publications included in my thesis, namely, Professor Ville V. Lehtola, Dr. Jouko Kalmari, Tuomo Palonen, and Mikko Vihlman. I especially wish to thank Professor Lehtola, who has helped me a lot by guiding me through the writing part of this thesis and also co-authored Publication I, and Dr. Jouko Kalmari, who also participated in several publications included in this thesis. Furthermore, I wish to thank Professor Ville Kyrki for assisting me with the sensor fusion

algorithms related to inertial measurements and Professor Themistoklis Charalambous for his support with particle filtering. I also wish to thank Matthew Billington for his excellent proofreading services.

The research for the thesis also included a large amount custom hardware, such as an instrumented prototype forest machine, rotating laser scanners and a camera lift mechanism. I wish to thank Matti Öhman for participating in planning and building the crane joint position sensor instrumentation for our prototype forest machine, the camera lift mechanism, and the two 3D laser scanner prototype constructions used in this thesis. I also wish to thank laboratory engineers Tapio Leppänen and Vesa Korhonen for helping with planning and building the equipment. Furthermore, I would like to thank Sami Kielosto for helping me with the electronics design. I also wish to thank Raimo Linkolehto from MTT Agricultural Engineering Research (at the Vakola premises, which was later merged into Natural Resources Institute Finland), who serviced our prototype forest machine for years and also helped with other hardware-related issues.

I have also been educated by the multitude of discussions with my colleagues during this long project. In that respect, I wish to thank Mikko Miettinen, Jakke Kulovesi, Teemu Tammi, Ville Matikainen, Andrei Sandru, Visa Jokelainen, Ville Toiviainen, Niko Nyrhilä, Juha Backman, and Timo Oksanen. In addition, I would like to thank the unit of Cognitive Science at the University of Helsinki, where I completed a bachelor's degree in cognitive science alongside my doctoral studies. I wish especially to thank Docent Otto Lappi for all his efforts to direct and teach his cognitive science students. Those studies have clearly guided my thinking towards perception being an active and highly complex process.

The research behind this thesis was performed between 2010 and 2018 with the Autonomous Systems research group at the department of Automation and Systems Technology and at the department of Electrical Engineering and Automation, Aalto University, and at the department of Remote Sensing and Photogrammetry at the Finnish Geospatial Research Institute (FGI), which is part of the National Land Survey of Finland. After that, I have worked at the FGI as a researcher and have consequently struggled to find time to write this summary.

I wish to thank all the funders who have made this project possible. The thesis was initially financed by the Doctoral School of the School of Electrical Engineering, Aalto University. The remainder of the funding for this prolonged work has come from several research projects, such as the Tekes (the Finnish Funding Agency for Technology and Innovation) NeoSilvix project, and two large Tekes SHOK (Strategic Centres for Science, Technology and Innovation) consortia: Value Through Intensive and Efficient Fibre Supply (EffFibre) and Data to Intelligence (D2I), where our research group worked on the Forest Big Data research area. In addition,

this work has been supported by the COMBAT/Pointcloud project[1], which is funded by the Strategic Research Council of the Academy of Finland.

In addition, I greatly appreciate the financial support provided by multiple foundations and funds, including the Aino and Kaarlo Tiisala Fund of the Satakuntalainen Osakunta student nation, the Finnish Foundation for Technology Promotion, and the Automation Foundation in Finland.

Furthermore, I would like to thank Research professor Harri Kaartinen, Professor Antero Kukko, and Professor Juha Hyyppä for supporting my work and allowing me to complete the thesis while working with the COMBAT/Pointcloud project[1]. The work is now continued in the Academy of Finland's Flagship of Science for Forest-Human-Machine-Interplay (UNITE)[2]

Finally, I would like to thank my family and friends for supporting me during this long process. Special thanks go to my partner Jenna Järvenpää, who has tolerated me writing this thesis during weekends and holidays.

Espoo, September 19, 2023,

Heikki Hyyti

[1]Read more about the COMBAT project at http://pointcloud.fi
[2]Read more about UNITE flagship at http://uniteflagship.fi

Contents

List of Publications

This thesis consists of an overview and of the following publications which are referred to in the text by their Roman numerals.

I Hyyti, Heikki, Lehtola, Ville V., and Visala, Arto. Forestry crane posture estimation with a two-dimensional laser scanner. *Journal of Field Robotics*, 25 pages, June 2018.

II Hyyti, Heikki and Visala, Arto. A DCM based attitude estimation algorithm for low-cost MEMS IMUs. *International Journal of Navigation and Observation*, 2015, Article ID 503814, 18 pages, July 2015.

III Kalmari, Jouko, Hyyti, Heikki, and Visala, Arto. Sway estimation using inertial measurement units for cranes with a rotating tool. In *the 2013 IFAC Intelligent Autonomous Vehicles Symposium*, Gold Coast, Australia, 6 pages, June 2013.

IV Palonen, Tuomo, Hyyti, Heikki, and Visala, Arto. Augmented reality in forest machine cabin. In *the 20th World Congress of the International Federation of Automatic Control (IFAC 2017)*, 8 pages, July 2017.

V Hyyti, Heikki, Kalmari, Jouko, and Visala, Arto. Real-time detection of young spruce using color and texture features on an autonomous forest machine. In *the 2013 International Joint Conference on Neural Networks (IJCNN)*, Dallas, TX, USA, 8 pages, August 2013.

VI Vihlman, Mikko, Hyyti, Heikki, Kalmari, Jouko, and Visala, Arto. Detection and species classification of young trees using machine perception for a semi-autonomous forest machine. In *2015 IEEE International Conference on Robotics and Automation (ICRA)*, Seattle, Washington, 6 pages, May 2015.

VII Hyyti, Heikki and Visala, Arto. Feature based modeling and mapping of tree trunks and natural terrain using 3D laser scanner measurement system. In *the 2013 IFAC Intelligent Autonomous Vehicles Symposium*, Gold Coast, Australia, 8 pages, June 2013.

Author's Contribution

Publication I: "Forestry crane posture estimation with a two-dimensional laser scanner"

The author designed and implemented the proposed crane posture estimation method and performed the experiments. The method was finalized and the journal article co-authored with Professor Ville Lehtola. The work was performed under the guidance of Professor Arto Visala.

Publication II: "A DCM based attitude estimation algorithm for low-cost MEMS IMUs"

The article was written, and its contents were solely based on, the work of the author. The work was performed under the guidance of Professor Arto Visala.

Publication III: "Sway estimation using inertial measurement units for cranes with a rotating tool"

The idea of estimating the tool rotation using two IMUs and a Kalman filter came from the author. Dr. Jouko Kalmari then derived the proposed sway estimation algorithm and held the main responsibility for writing the conference paper. The author designed and implemented the hardware and micro controller software for the proposed sway estimation method to use it in real time and helped write the conference paper. The experiments were performed jointly by the author and Dr. Jouko Kalmari. The work was performed under the guidance of Professor Arto Visala.

Publication IV: "Augmented reality in forest machine cabin"

The author proposed the idea of implementing the first augmented reality demonstration in a forest machine cabin and participated in implementing

the algorithm and writing the conference paper. Tuomo Palonen, as part of his master's thesis, by using the software provided by the author, held the main responsibility for implementing the marker detection algorithm in the software, and designing the proposed sensor fusion algorithm to fuse a rotation and position estimate from the marker detection algorithm with a separate attitude estimate acquired from inertial measurements. The experiments were conducted and the software created jointly by the author and Tuomo Palonen. The work was performed under the guidance of Professor Arto Visala.

Publication V: "Real-time detection of young spruce using color and texture features on an autonomous forest machine"

The conference paper was jointly written by the author and Dr. Jouko Kalmari. The proposed color transformation and k-NN voting scheme for detecting spruce location were designed by the author. The proposed texture analysis method for detecting young spruce was designed and implemented by Dr. Jouko Kalmari. The experiments were conducted and the software created jointly by the author and Dr. Jouko Kalmari. The work was performed under the guidance of Professor Arto Visala.

Publication VI: "Detection and species classification of young trees using machine perception for a semi-autonomous forest machine"

The author implemented the 3D scanning software, planned the sensor setup, and installed the sensors. The experiments were conducted jointly by the author and Mikko Vihlman. The classification method was developed and the conference paper mainly written by Mikko Vihlman with the help of other authors. The work was performed under the guidance of Professor Arto Visala.

Publication VII: "Feature based modeling and mapping of tree trunks and natural terrain using 3D laser scanner measurement system"

The conference paper was written, and its contents were solely based on, the work of the author. The work was performed under the guidance of Professor Arto Visala.

List of Figures

List of Tables

Nomenclature

Indexing and voting

i, j	indices
k	a time index
l	a laser range measurement index
L	a label
V_L	number of votes for the label L
N_b	number of histogram bins
N_c	number of cells
N_p	number of particles
s_i	an ith sensor

Parameters and measures

a_x	acceleration in x axis (m/s^2)
b_α	a friction parameter for α ($\frac{\mathrm{Nm}}{\mathrm{kg\ rad/s}}$)
b_x	gyroscope bias around x axis (rad/s)
C_l	a count-to-the-middle measure for lth laser range
d	diameter (m)
d_4	an extension length of a forestry crane (m)
$\dot{d}_4$	an extension velocity (m/s)
$D_i(k)$	a kernel density estimate around the ith particle at time index k
$\hat{D}(k)$	a value of the MAP estimate of kernel density estimates
g	magnitude of Earth's gravity (m/s^2)
l_1, l_2, l_3	characteristic lengths (m)
N_l	a cluster-size measure for lth laser range
q_L	a quality threshold for the label L
r	range (m)

r_l	lth range measurement in a laser scan (m)
t	time (s)
v	speed (m/s)
w_i	a weight of the ith particle
w_i^-	a predicted weight of the ith particle
W_i	a normalized weight of the ith particle
α, β, γ	general symbols for angular values (rad)
$\Delta\beta$	a resolution of the laser scanner (rad)
$\Delta t(k)$	a sample period, time between indices $k-1$ and k (s)
ϵ_d	a range detection margin (m)
η	a normalization constant
$\theta_1, \theta_2, \theta_3$	crane joint angles for slew, lift, and transfer (rad)
$\dot{\theta}_1, \dot{\theta}_2, \dot{\theta}_3$	angular velocities of the crane joint angles (rad/s)
ν	a process noise
σ_A, σ_A^2	a standard deviation and a variance of A
υ	a measurement noise
ϕ_l	an angle of lth range measurement in a laser scan (rad)
ψ, θ, ϕ	Euler angles yaw, pitch, and roll in ZYX convention (rad)
ψ_s, ψ_c	abbreviations for $\sin(\psi)$ and $\cos(\psi)$
$\dot{\psi}, \ddot{\psi}$	an angular velocity and an angular acceleration of ψ
ω_x	an angular velocity around x axis (rad/s)
Ω	a solid angle (sr)

Distributions and probabilities

$\mathscr{L}(A)$	likelihood of A
$L(A \mid B)$	a one-dimensional fitness function (approx. likelihood) of A given B
$\mathscr{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	a joint normal distribution with a mean $\boldsymbol{\mu}$ and a covariance $\boldsymbol{\Sigma}$
$p(A)$	probability of A
$p(A \mid B, C)$	conditional probability of A given B and C
$\mathscr{U}(0,1)$	a standard uniform distribution between 0 and 1

Vectors

$\mathbf{a}_k$	a (non-gravitational) acceleration at time index k
$\mathbf{f}$	a vector of features F_1, F_2, …
$\mathbf{p}_k$	a parameter vector at time index k
$\mathbf{u}$	a control input, i.e. a vector of control values $u_1, u_2, \ldots, u_{N_u}$
$\hat{\mathbf{v}}_{joint}(k)$	an estimated joint velocity vector at time index k

$\mathbf{x}$	a system state, i.e., a vector of state variables $x_1, x_2, \ldots, x_{N_x}$
$\mathbf{x}'$	an augmented state
$\hat{\mathbf{x}}_k$	a system state estimate at time index k
$\hat{\mathbf{x}}_k^-$	a predicted system state estimate at time index k
$\mathbf{x}_i(k)$	an ith particle at time index k
$\mathbf{x}_i^-(k)$	an ith predicted particle at time index k
$\mathbf{y}$	a measurement, i.e., a vector of measured values
$\hat{\mathbf{y}}_k^-$	a predicted measurement estimate at time index k

Functions and matrices

$\mathbf{1}_i$	an indicator function for particle i
${}^n_b\mathbf{C}$	a DCM from the body-fixed frame to the navigation frame
${}^n_b\mathbf{C}_i$	ith row vector in a DCM matrix
${}^n_b C_{ij}$	a value at the ith row and jth column in a DCM matrix
$f_k(\cdot)$	a time-varying nonlinear system equation at time index k
$h_k(\cdot)$	a time-varying nonlinear measurement equation at k
$[\mathbf{v}\times]$	a skew-symmetric matrix of a vector $\mathbf{v}$
$\mathbf{0}_{3\times 3}$	a 3×3 sized matrix of zeros
$\mathbf{I}_3$	a 3×3 sized identity matrix
$\mathbf{F}_k, \mathbf{G}_k$	matrices representing linear dynamic system model at k
$\mathbf{H}_k$	a linear measurement model at time index k
$\mathbf{K}_k$	a Kalman gain at time index k
$\mathbf{L}_k, \mathbf{M}_k$	matrices representing linearized noise processes at k
$\mathbf{P}_k$	a covariance of the state $\mathbf{x}_k$ at time index k
$\mathbf{P}_k^-$	a predicted covariance of the state $\mathbf{x}_k$ at time index k
$\mathbf{Q}$	a state-prediction covariance of a noise process $\{\boldsymbol{v}\}$
$\mathbf{R}$	a measurement covariance of a noise process $\{\boldsymbol{\upsilon}\}$

Sets

$\mathbf{S}$	a set of system states
$\mathbf{S_f}$	a subset of feasible states ($\mathbf{S_f} \subset \mathbf{S}$)
$\mathbf{T}$	a set of targets
$\mathbf{U}$	a set of control inputs $\mathbf{u}$
$\mathbf{X}(k)$	a set of particles $\mathbf{x}_i, i \in [1, \ldots, N_p]$ at time index k
$\mathbf{Y}$	a set of measurements $\mathbf{y}$

Color channels

EG,RB,I	excessive green, redness-blueness, and intensity
R,G,B	red, green, and blue

Abbreviations

2D	two-dimensional
3D	three-dimensional
ABA	area-based approach
ACF	autocorrelation function
AHRS	attitude and heading reference system
AI	artificial intelligence
ALS	airborne laser scanning
BPF	bootstrap particle filter
CAN	controller area network
CTL	cut-to-length logging
CUDA	Compute Unified Device Architecture
DCM	direction cosine matrix
DGNSS	differential global navigation satellite system
DGPS	differential Global Positioning System
DWT	discrete wavelet transform
EKF	extended Kalman filter
FoV	field of view
FPS	frames per second
GDP	gross domestic product
GLCM	gray level co-occurrence matrix
GLONASS	Globalnaya navigatsionnaya sputnikovaya sistema[3]
GNSS	global navigation satellite system
GPS	Global Positioning System
GPU	graphics processing unit
HMD	head-mounted display
HSV/HSI/HSL	hue, saturation, lightness value / intensity / luminance
ICP	iterative closest point
IMU	inertial measurement unit
INS	inertial navigation system
KF	Kalman filter

[3]transliteration from Russian, Глобальная навигационная спутниковая система

LBP	local binary pattern
LIDAR	light detection and ranging
LIO	lidar-inertial odometry
LOAM	lidar odometry and mapping
MARG	magnetic angular rate and gravity
MEMS	micro-electro-mechanical system
MLS	mobile laser scanning
PDF	probability density function
PF	particle filter
RF	radio frequency
RFID	radio frequency identification
RGB	red, green, and blue
RGB-D	red, green, blue, and depth
RIO	radar-inertial odometry
RMSE	root mean square error
RPM	revolutions per minute
SAR	synthetic-aperture radar
SAS	synthetic-aperture sonar
SIR	sampling importance resampling
SIS	sequential importance sampling
SLAM	simultaneous localization and mapping
TLS	terrestrial laser scanning
ToF	time of flight
UAV	unmanned aerial vehicle
UKF	unscented Kalman filter
UWB	ultra wide band
VIO	visual-inertial odometry

1. Introduction

Boreal forests, which extend as a band through Fennoscandia, Russia, Alaska, and Canada, cover an area of about 1.7 billion hectares (Vanhanen, Jonsson, Gerasimov, Krankina, & Messieur, 2012). They are globally significant, since they produce approximately 45% of the world's stock of growing timber. These forests are also of local importance, since the forestry sector (including the sub-sectors logging, solid wood products, pulp & paper, and furniture) directly accounts for approximately 1% of gross domestic product(GDP)in Russia, Norway and Canada, 2% in Sweden, and 3% in Finland (Y. Li, Mei, & Linhares-Juvenal, 2019; World Bank, 2019). Moreover, the importance of boreal forests is expected to increase in the future. For example, it has been estimated that Finland's annual timber harvest volume could be increased by more than 50%, from approximately 55 million to more than 85 million m^3 by using more intensive forest management practices (FIBIC, 2014). Similarly, in Sweden, forest production could be almost doubled by using more intensive forest management methods (Nilsson, Fahlvik, Johansson, Lundström, & Rosvall, 2011).

There is an urgent need to increase the productivity[4] of the forest sector. The European Commission has set a target to raise the share of renewables to at least 32% of energy consumption by 2030 (Council of European Union, 2018), and there are plans to increase this target to 42.5% (Wilson, 2021). Forests are considered an important resource to meet these renewable energy targets. Errera, Dias, Maya, and Lora (2023) forecast in their optimistic, yet plausible scenario, that bioenergy share in the global energy matrix could increase from today's 9.8% to 37%, but it would require gains in productivity. However, it should be noted that an increase in harvest demand of even 20% has been estimated to approach the maximum harvest potential of the world's forest area (Pilli, Grassi, Kurz, Fiorese, & Cescatti, 2017).

Much has already been achieved to increase productivity. For example, according to Nordfjell, Björheden, Thor, and Wästerlund (2010), the pro-

[4]Productivity can be understood as the total efficiency of resource (e.g., money, work, and material) usage per number of units produced (Schreyer & Pilat, 2001).

ductivity of logging by the Swedish forest enterprise SCA increased almost three-fold in the period 1985—2010. Nonetheless, such increases have been driven primarily by the mechanization of forest machinery. Mechanization has a long history, from manually operated tools through powered tools towards increasingly automated forest machinery (Silversides & Rajala, 1997; Lindroos, La Hera, & Häggström, 2017). However, today, the easy mechanical solutions for increasing the productivity of machinery have already been attempted. Moreover, further increasing the size, power, or capabilities of forest machinery is expensive and might not bring the expected benefits. According to Lindroos et al. (2017), there is no single perfect solution in view, but, instead, future forest machine development will be influenced by the necessity for local adaptation.

Some tasks are more difficult to mechanize and automate than others. In many countries, due to rising labor costs, full mechanization of the harvesting and transport system has already become competitive (Asikainen, Anttila, Verkerk, Diaz, & Röser, 2011). However, in these countries, many early forest-management operations are still mostly performed either manually or motor-manually (e.g., Kärhä et al., 2014; Ersson, 2014; Strandström, 2016). Manual work is favored in these tasks, since mechanized techniques either struggle to achieve the required quality or are insufficiently cost efficient (Pettersson, Fahlvik, Karlsson, & och Skogsstyrelsen, 2012; Hämäläinen et al., 2013; Uotila, 2017). Management of younger stands remains a particular challenge. Young trees are smaller and grow more densely, which impedes the work of forest machine operators using crane-mounted tools. For example, when thinning or cleaning a young stand, the operator should avoid damaging the remaining target trees while cutting or uprooting neighboring ones. This problem is exacerbated by the fact that young target trees are usually concealed by other faster growing vegetation. Thus, in many cases, laborious motor-manual work with a clearing saw is still preferred.

Productivity increases are also limited by the capabilities of machine operators, who are under high pressure to work efficiently with limited information. In most cases, they must rely on their own senses, which is difficult in poor lighting conditions, especially during night shifts, rain, or snowfall. Moreover, it should be noted that, even in good weather, only one side of each tree can be inspected from the cabin (Liziniewicz, Ekö, & Klang, 2016). It has been found that the complexity of modern forestry work calls for long training before the operator can reach full productivity (Asikainen et al., 2011). In addition, the differences between machine operators in productivity are significant (Ovaskainen, 2009; F. T. Purfürst & Erler, 2011), which suggests that the operator's skills or capabilities are the bottle neck.

Consequently, instead of developing traditional mechanized solutions where the operator performs all the sensing and machine control, produc-

tivity increases are currently sought by adding more sensors, algorithms, and artificial intelligence to forest machinery. Research is ongoing in areas such as operator tutoring and decision support systems (e.g., Väätäinen et al., 2011; Mohtashami, Bergkvist, Löfgren, & Berg, 2012; Magliocchetti, Prandi, Panizzoni, Lotto, & De Amicis, 2015), mixed, virtual, and augmented reality displays (e.g., Nordlie & Till, 2015; Palonen, 2016), intelligent boom control and automated functions (e.g., Hansson & Servin, 2010; Ortiz Morales et al., 2014; Kalmari, Backman, & Visala, 2014), tele-operated forestry vehicles (e.g., Milne, Chen, Hann, & Parker, 2013; Westerberg & Shiriaev, 2013), unmanned self-navigating vehicles (e.g., Vestlund & Hellström, 2006; Hellström, Lärkeryd, Nordfjell, & Ringdahl, 2009; Ringdahl et al., 2011), forest mapping for inventory purposes (e.g., Melkas, Miettinen, Hämäläinen, & Einola, 2014; S. W. Chen et al., 2020), and unmanned aerial vehicles (UAV) in forestry (e.g., Torresan et al., 2017).

It is believed that, in the future, better use of information (e.g., by using sensors, algorithms, and optimizers) in forest machinery will help increase productivity. Vanclay (2011) proposes that sensing technology will allow harvesters to optimize vehicle movements and improve the handling of harvested material in the future. In addition, harvesters could provide a comprehensive inventory of the residual stand and the soil underneath, which may assist the later management of the residual forest.

In this work, perception systems are proposed for autonomous or semi-autonomous forest machinery to tackle the many challenges related to gaining the required productivity increases. By enabling the effective use of various sensors in a forest machine, productivity could be improved by 1) finding novel solutions to enable the use of a minimal number of low-cost sensors to provide maximal information, 2) improving the user interface of the forest machine, thus allowing the machine operator to make better decisions while working in the forest, 3) increasing the autonomy of the forest machine in selected tasks to offer more time for the operator to focus on planning and decision making, 4) providing more accurate measurements from the forest machine itself and from the surrounding forest to enable accurate data collection on the forest machine operations and the work quantity and quality, and 5) enabling the collection of forest inventory information from the residual forest after operation.

This work aims to find solutions, applicable for industrial purposes, for placing suitable sensors on forest machinery and to develop methods to process the sensor data to produce a more usable form of knowledge about the machine and its surroundings. The work contains several model-based and sensor-fusion methods for integrating measurements into an artificial understanding in real time—a machine perception of the vehicle itself and its environment.

1.1 Background

The design of a perception system for autonomous forest machinery is dependent on many aspects of the environment where the machine operates. Traditionally, to increase the level of automation, for instance in factories, the common approach has been to simplify the problem by simplifying the environment. To accomplish this, the environment has been modified until it corresponds to an applicable mathematical model (Bessière, Laugier, & Siegwart, 2008, p. 4). For example, in factories where a robot picks items from a conveyor belt, the machine vision problem has been simplified by using specially designed lighting, and the belt in the background has a matte, untextured surface finish to avoid reflections and to expose target objects on the belt (e.g., Horn, 1986, pp. 3–4).

The forest environment, however, is not a factory that can be modified to ease the task of the forest machine. Forests are living, naturally growing ecosystems and home to a large variety of plants, animals, fungi, and bacteria. Hence, a forest is defined as an *uncontrolled environment*[5] in contrast to a factory, which is an example of a *controlled environment*. Therefore, perception problems are also different. In a controlled environment, the prior model of the environment can be trusted. By contrast, in an uncontrolled environment, all methods must be devised to take account of the high level of uncertainty that arises from a lack of knowledge (Bessière et al., 2008, p. 4). In an uncontrolled environment, the methods must be robust against random disturbances, and they should be able to function with approximate models and imperfect observations.

The forest as an uncontrolled environment raises some new challenges compared to the factory environment. An autonomous forest machine requires, for example, the ability to measure the trees, select suitable trees for cutting, and estimate traversability in the forest. Traversability (e.g., Suger, Steder, & Burgard, 2015; Ahtiainen, Stoyanov, & Saarinen, 2017; Ruetz, Borges, Suenderhauf, Hernández, & Peynot, 2022) relates to the robot's ability to safely navigate and distinguish obstacles such as fallen trees, rocks, potholes, cliffs, and ravines from traversable ground. Furthermore, wet areas and, especially in winter, thin ice on top of them can be a potential hazard (Vestlund & Hellström, 2006).

In the following sections, the relevant aspects of boreal forests and the Nordic countries as an environment are first introduced. Then, the current available information sources from forests are outlined. These are forest inventory systems and the sensors mounted on forest machinery. Finally, the basics of machine perception methods and autonomous robots are introduced.

[5] Some authors define a forest as highly unstructured environment (e.g., Lindroos, Mendoza Trejo, La Hera, & Ortiz Morales, 2019), but since nature is full of various biological structures, the level of human control is emphasized here instead.

Fennoscandian Boreal Forests as an Environment

Boreal forests grow on gently rolling terrain, either on glacial till, or on shallow-soiled and infertile uplands alternating with wetlands and poorly drained organic soils (Burton et al., 2003). The climate and soils are cold, which causes trees to grow slowly—mostly during short growing seasons. These forests are dominated by a few tree species, mostly conifer species of the pine (*Pinus*), larch (*Larix*), spruce (*Picea*), and fir (*Abies*) genera and broadleaf species of usually the birch (*Betula*), poplar (*Populus*), willow (*Salix*), alder (*Alnus*), and rowan (*Sorbus*) genera (Burton et al., 2003).

The intensity of forest activities varies significantly across the boreal zone. It ranges from timber logging with minimal consideration for forest regeneration, through extensive management with simple silvicultural approaches, to extremely intensive management with frequent management interventions (Vanhanen et al., 2012). The least managed forests are found in the Russian Federation (e.g., only one third of the stands managed are thinned), and the most intensively managed forests are in Fennoscandia (e.g., much land is drained, the forest road network is dense, and most of the stands are thinned at least once) (Burton et al., 2003). This is important, since the level of structure of the otherwise uncontrolled environment usually increases when the forests are more managed.

The ownership structure is also strikingly different around the boreal zone. Forests are mainly owned by public (regional, provincial, and federal) institutions in Russia, Alaska, and Canada, and by private (industrial or non-industrial) entities or in Fennoscandia (Vanhanen et al., 2012). The large institutional ownership seen in Russia, Alaska, and Canada is a legacy of the extensive harvesting rights assigned to primary processing facilities (e.g., pulp mills or sawmills) (Burton et al., 2003). By contrast, there are over 600,000 forest owners in Finland, and the size of the average forest holding is only 25 hectares, with an average stand size of less than 2 ha (Kankare et al., 2017). These varying ownership characteristics have different practical implications for forestry around the area.

Forestry in the Nordic countries has long been based upon the principle of sustainable management (NOLTFOX, 2006) and Nordic countries have been shown to succeed in keeping the forest sector sustainable (A. C. Hansen, Clarke, & Hegnes, 2021). In intensively managed forests in Fennoscandia, guided by official recommendations, forest owners mainly focus on only a small number of commercially viable tree species. In recent decades, conifers have been favored and broad-leaved trees have usually only been left in gaps or as replacements for severely damaged conifers (Fahlvik, Ekö, & Petersson, 2015). For example, in Finland, as a result of regulations, recommendations, and intensive management, roughly 50% of the growing stock is Scots pine (*Pinus sylvestris*), 30% Norway spruce (*Picea abies*), and 17% silver and downy birch (*Betula pendula* and *Be-*

tula pubescens), while only 3% of the trees are other hardwoods, mainly gray alder (*Alnus incana*) and aspen (*Populus tremula*) (Tapio, 2014; Lier, Korhonen, Tuomainen, Viitanen, & Mutanen, 2017).

Due to the long tradition of intensive forest management, most forests in Fennoscandia are even-aged, same-species forests. However, some forests contain mixed species growing either at the same age or in multiple age groups. This dualism relates to two conventional management systems, even-aged management with a clear-cutting system and uneven-aged management employing a single tree-selection system. These represent the two extreme cases in the continuum of removed versus retained trees per each cutting event (Kuuluvainen, Tahvonen, & Aakala, 2012). Between these two extremes, a wide range of methods have been proposed in which the forest cover remains more or less continuous, although sometimes patchy and partly open to facilitate regeneration (Kuuluvainen et al., 2012).

(a) Mature even-aged Scots pine forest

(b) Mature even-aged Norway spruce forest

(c) Young even-aged Scots pine forest

(d) Uneven-aged mixed species forest

Figure 1.1. Examples of Finnish forests visited in the research for Publication VII

Because even the most managed man-made forests are composed of native species, they differ only slightly from naturally regenerated but intensively managed stands (Hakkila, 1989). However, for detecting and measuring trees, these managed forests are significantly different from forests in a natural state. They are somewhat controlled environments, but the level of structure varies significantly. This can be seen from the set of examples of several typical Finnish forests in Figure 1.1. As can also be noted from the example figures, the difficulty of traversability estimation (i.e., passable ground and obstacle detection), tree detection, species classification, and tree measurement tasks may vary from reasonably easy to extremely difficult in Fennoscandian forests.

Other Specific Aspects of the Nordics

The Nordic countries Sweden, Finland, and Norway are among the most important producers of wood and forest products in the world (NOLTFOX, 2006). The northern forest industry is currently dominated by a few large global companies (Donner-Amnell, Lehtinen, & Sæther, 2017). This increased international orientation and ownership has changed how companies view their stake holders and, for example, how they respond to environmental questions and aim for social responsibility (Donner-Amnell et al., 2017). Today, the main challenge for the forestry sector in the Nordic countries is to maintain their global competitiveness while balancing production costs with environmental objectives (Vanhanen et al., 2012). One plausible way to couple profitable but environmentally friendly and socially acceptable operation is through technological development. For example, the Finnish forest industry expects next-generation forest inventory techniques to improve current wood procurement practices (Holopainen, Vastaranta, & Hyyppä, 2014).

Finland and Sweden are home to many competitive forest machine manufacturers, providing the most technologically advanced cut-to-length forest operations technology (Nordfjell et al., 2010; Lindroos et al., 2017), In addition, the Nordic countries are world leaders in the utilization of forest biomass for energy production (Routa, Asikainen, Björheden, Laitila, & Röser, 2013) and host several significantly large forestry sector companies (Blocker, Bromley, & Murdoch, 2016). Moreover, the sector enjoys a high level of research and development investment from governments and the private sector alike. This is important, since developing a technology as complex as autonomous forest machinery requires a significant amount of research and investment (Lindroos et al., 2019).

Forest Inventories

Forest inventories employing aerial and satellite imagery and airborne laser scanning (ALS) have long been in operation in the Nordic countries (Næsset et al., 2004; Kangas et al., 2018). Currently, forest inventory attributes are commonly derived using an area-based approach (ABA) in which low-density (~0.5 pulses per m^2) ALS data are used to generalize field-measured inventory attributes over the entire inventory area (Holopainen et al., 2014). However, the use of higher density (~5 pulses per m^2) ALS data was introduced during 2020 in Finland (Laaksonen, 2019). Nonetheless, remote sensing methods are often used merely as background material for manual generation of forest inventories (Kangas et al., 2018). Nevertheless, there is plenty of potential in forest data. Owners' associations can use them to plan individual operations, and forest companies can utilize them to optimize the bucking of logs, timber trade, and plan wood procurement, harvests, and logistics, to mention but a few benefits (Kangas et al., 2018). It has been estimated that better forest data could increase the profitability of the forestry sector by more than 250 million euros a year in Finland (Kangas et al., 2019).

Information on forest resource attributes such as species-specific timber assortments, the diameter distribution of trees, tree quality, and biomass distribution cannot be obtained accurately enough from the current inventory system (Kankare et al., 2017). Therefore, there is a strong incentive to develop more accurate data collection methods and forest inventory techniques. Single-tree-level forest inventory methods have been proposed to solve the problem (Holopainen et al., 2014; Kankare et al., 2017). However, they require significantly denser point clouds than traditional statistical ABA methods. Furthermore, as each tree is measured and modeled individually, the tree trunk (or at least the stem and the largest branches) should also be incorporated into the measurements. This limits the use of aerial imagery and ALS data and favors the use of terrestrial and mobile laser scanning approaches. In addition, data measured by harvesters offer much potential for single-tree-level methods (Olivera & Visser, 2016; Kankare et al., 2017).

Modern Forest Machinery

Nordic cut-to-length (CTL) harvesting using two specialized machines, a harvester and a forwarder, is the most technologically advanced forest harvesting scheme in the world (Nordfjell et al., 2010; Lindroos et al., 2017). In CTL, forest machine operators use a harvester to fell, delimb, and cross-cut the tree into logs, and a forwarder to transport the logs to a roadside landing (Gellerstedt & Dahlin, 1999; Häggström, 2015). The main tools of these machines are large hydraulic serial manipulators, known

as cranes. To control these cranes, as well as the machine, operators sit inside the cabins and use joysticks to control the hydraulic valves (La Hera & Morales, 2019). Modern forest harvesters allow the boom tip to be controlled in Cartesian coordinates, which removes the need to control each hydraulic valve individually (La Hera, Morales, & Mendoza-Trejo, 2021). The machine operator is responsible for planning the operations and for performing synchronized control actions to safely work in the forest.

One of the features of the CTL system is that in the first thinning phase, harvesting tracks, called strip roads (i.e., treeless corridors for the machine to travel around the forest), must be cut at about 20 meters distance from each other, since the harvester crane reach is approximately 10 meters (Gellerstedt & Dahlin, 1999; Ovaskainen, Uusitalo, & Sassi, 2006). Operators are free to choose the location of the strip roads as well as the selection of thinned trees in their normal work (Ovaskainen, 2009). In the past, tree marking prior to harvesting was a common task performed by experienced foresters, but currently prior selection is omitted to reduce costs (Holzleitner et al., 2019). These strip roads are used for forwarding and on later thinnings, and they exert a large effect on the productivity of later operations.

CTL harvesting is challenging for machine operators. It has been described as joystick-intense, mentally demanding work in which visual information and supervision play a key role (Gellerstedt, 2002; Häggström, 2015). The operator requires silvicultural knowledge and experience with different kinds of stands, thinning principles, seasons and weather conditions. For example, it takes an average of five years to become fully skilled in thinning (Gellerstedt, 2002). Productivity differences between operators of over 40% have been measured (e.g., Ovaskainen, Uusitalo, & Väätäinen, 2004).

Modern commercial computerized CTL machines contain many built-in sensors, and many new sensors have been proposed in the research literature. Such sensors may be divided into three distinct categories: 1) linear and angular position sensors, 2) orientation and global positioning sensors, such as inertial measurement units and satellite navigation sensors, and 3) remote sensing sensors, such as lidars[6], radars, sonars, and optical cameras (Lindroos, Ringdahl, La Hera, Hohnloser, & Hellström, 2015).

In CTL machinery, linear and angular position sensors are used, for example, to measure the posture of the forestry crane (Lindroos et al., 2015) as well as log length and diameter (Miettinen, Kulovesi, Kalmari, & Visala, 2010) and to estimate traveled path and distance using wheel odometry (Hellström et al., 2009). In turn, inertial techniques and tilt sensors are used to measure the inclination of the machine or crane parts (Lindroos et al., 2015). Furthermore, it has been proposed that the position

[6]Lidars which have also been called laser scanners, scan the environment by combining multiple point-wise range measurements as a point cloud.

of the tip of the forestry crane be estimated and combined from multiple inclination measurements (e.g., Vihonen, Honkakorpi, Tuominen, Mattila, & Visa, 2016).

Satellite navigation systems are commonly used to pinpoint the machine on the map for the operator. However, in a forest, positioning accuracy is a limiting factor. Positioning errors under the forest canopy are usually larger than 4 meters with conventional satellite navigation sensors and about 0.7 meters with a high-end inertial-navigation integrated satellite navigation system (Kaartinen et al., 2015). Using modern smartphones Tomaštík, Chudá, Tunák, Chudỳ, and Kardoš (2021) achieved centimetre-level accuracy under open-area conditions but, in forest, the accuracies varied from meters to tens of meters. In a recent evaluation by T. Purfürst (2022), the most accurate smartphone with multi-frequency GNSS receiver reached 3.2 m average positioning error under the forest canopy.

The use of remote sensing instruments has often been proposed in the research literature. These sensors mostly concern 2D laser scanners (e.g., Miettinen, Öhman, Visala, & Forsman, 2007; Öhman et al., 2008; Rossmann et al., 2011; Zheng, Liu, Wang, & Yang, 2012; Salmivaara et al., 2018), but 3D lidars (e.g., Sihvo, Virjonen, Nevalainen, & Heikkonen, 2018; Pierzchała, Giguère, & Astrup, 2018), machine vision solutions (e.g., Kulovesi, 2009; Miettinen et al., 2010; Kalmari, Kulovesi, & Visala, 2011), and combined sensor fusion solutions (e.g., D. Wang, Lia, Wang, & Li, 2013) have also been proposed. However, they are not yet used in commercial forest machinery (Lindroos et al., 2019). Only some rear-view and surveillance cameras have been installed in commercial CTL machines to assist the operator (e.g., Komatsu, 2019).

Even without remote sensing, forest machinery can nonetheless collect forest inventory data. If properly calibrated, the harvester head can individually measure the dimensions of harvested logs using its mechanical sensors (Rasinmäki & Melkas, 2005; Marshall, Murphy, & Boston, 2006). If the position of the harvester head is also measured (Lindroos et al., 2015), forest data may be composed from the information collected (Olivera & Visser, 2016). However, accurate harvester head positioning is difficult and unavailable in most commercial forest machines (Lindroos et al., 2015; Hauglin et al., 2017; Noordermeer, Sørngård, Astrup, Næsset, & Gobakken, 2021), and thus the use of harvester data is currently limited. In a recent work by Noordermeer et al. (2021), a forest machine was upgraded with a dual-wavelength high-end GNSS receiver and the harvester head was positioned with errors ranging from 0.14 to 2.85 m, with a mean of 0.88 m.

On the other hand, remote sensing with lidars, radars, and cameras are commonly used to collect data from various environments, including forests (C. Toth & Józków, 2016). Similarly, in robotics, for example in advanced driver assistance systems, cameras, radars, sonars, and lidars already play an important role (Ziebinski, Cupek, Erdogan, & Waechter, 2016).

These sensors offer much potential for collecting inventory information and guiding the machine operator when installed in commercial forest machinery. However, machine manufacturing companies must see a clear benefit to adding such sensors before they are willing to integrate them into their machinery. Another limiting factor is automatic sensor data processing. If remote sensing equipment are integrated into forest machinery, the data must be processed automatically with a minimal amount of human intervention to increase productivity. Such automatic sensor data processing is studied in the field of machine perception.

Machine Perception

According to Nevatia (1982), the field of machine perception concerns the building of machines that sense and interpret their environments. In robotics, the term perception is used to describe a system which allows a robot to perceive, comprehend, and reason about the surrounding environment (Premebida, Ambrus, & Marton, 2018). Machine (or robotic) perception is a way to connect the computer to the raw, "unwashed" world, and it is a key subfield of artificial intelligence (AI) (Russell & Norvig, 2016, p. 928). Perception enables a computer or robot to use its sensory input to gather information and present it in a way that is appropriate to the task at hand.

This thesis argues that perception should be built using sensors from multiple modalities. In addition to visual remote sensors such as cameras and lidars, the robot's perception capabilities can be improved by measuring the position and orientation of the robot and its parts. However, most earlier research on machine perception has focused on visual sensors, such as optical cameras (e.g., Nevatia, 1982). For these sensors, such research has, arguably, been conducted under the names of machine vision, computer vision (Jain, Kasturi, & Schunck, 1995), and photogrammetry (Granshaw & Fraser, 2015).

Similar to the way animals benefit from senses other than the visual, robots can also use other types of sensors and related perception methods to function better in the environment. Machine hearing (Lyon, 2010), also known as machine listening or computer audition (W. Wang, 2010), and machine touch, which concerns robotic tactile perception and surface properties (Edwards, 2013) belong to the field of machine perception. In addition, the proprioceptive senses of robots are studied within the field of machine perception. Such senses are, for example, vestibular sense for balance and kinesthetic perception for the robot's own motion (J. Ferreira, Lobo, Bessiere, Castelo-Branco, & Dias, 2013; S. Chaudhuri & Bhardwaj, 2018, p. 7). These other perception modalities can provide the system with new information that would otherwise be difficult or even impossible to obtain using visual sensors alone.

Early approaches to machine perception have largely relied on the recovery and manipulation of geometric world models (Nevatia, 1982; Elfes, 1989). At a low level, this has meant the extraction of geometric features such as line segments or surface patches from sensor data. From this *geometric* perspective, high-level sensing processes have been designed to use symbolic models, geometric templates, and heuristic prior assumptions of the environment. Elfes (1989) has referred to this as the geometric paradigm in robot perception.

Geometric approaches to machine perception succeed in highly controlled environments such as factory settings, but have limited applicability in more complex scenarios (Elfes, 1989). In uncontrolled environments such as forests, sensor measurements are inherently uncertain. This uncertainty arises from sensor limitations, noise, and the unpredictability of the uncontrolled environment (Thrun, 2000). In contrast to the geometric approach, less deterministic, probabilistic methods compute a probability distribution over what might be the situation in the world. Instead of relying solely on a single best guess (e.g., when a geometric shape is detected from the data), the probabilistic system attempts to make inferences from multiple observations based on likelihoods from the observations. As a result, a probabilistic robot may recover from errors, handle ambiguities, and integrate sensor readings in a consistent way. Furthermore, a probabilistic robot possesses a measure of its own ignorance, which is a key prerequisite for autonomous operation (Thrun, 2000).

The benefits of the probabilistic paradigm also involve trade-offs. The limitations of probabilistic algorithms are computational inefficiency and the need to approximate (Thrun, 2000). Probabilistic algorithms are inherently less efficient than their non-probabilistic counterparts because they consider entire probability densities instead of the best guess alone. In addition, most real environments are difficult to model, since they are continuous spaces, where algorithms assuming discrete posterior distributions are difficult to model and computationally infeasible (Thrun, Burgard, & Fox, 2005, pp. 5–6). This leads to the need for approximations in the models and the computation. However, research has led to a range of ideas and algorithms that enable the development of computationally feasible and sufficiently accurate methods (Thrun, 2000).

Autonomous Forest Machinery

Autonomous systems are defined by their ability to perform certain functions without the direct control of a human operator (Visser & Obi, 2021). According to Visser and Obi (2021), the challenges of forest operations are so complex that no fully autonomous systems currently work in timber harvesting. Previous research has suggested that the path towards fully autonomous forest robots will first involve a hybrid approach: semi-

autonomy, in which the robot cooperates with the human operator (e.g., Westerberg, 2014). Such cooperation may occur on site from the cockpit, by remote control from nearby, or from a greater distance by using, for instance, virtual-environment-based teleoperation (Westerberg & Shiriaev, 2013). Semi-autonomy can be considered a set of intermediate steps between the current mechanized CTL harvesting scheme (Gellerstedt & Dahlin, 1999) and full autonomous operation (Lindroos et al., 2017). This cooperation may be beneficial when the robot can outperform the human operator in some specific tasks, but, at the same time, some other tasks may be too difficult to automate. Hellström et al. (2009) lists three possible scenarios for increasing automation in forest machinery: 1) remote supervision of a semi-autonomous system where a remotely working operator may take control when necessary, 2) remote-control of a semi-autonomous harvester from a forwarder, and 3) use of semi-autonomous forwarders, called wood shuttles, which are remotely operated from the harvester for loading and unloading.

For autonomous or semi-autonomous operation in a forest environment, the robustness of machine perception is essential. Both fully autonomous and semi-autonomous systems require knowledge from the environment and from the robot's states, for example the pose of the robot and its crane. A probabilistic framework may enable the system to collect and fuse information from multiple uncertain sources and build a coherent view of the environment (Thrun et al., 2005, pp. 230–231). This could be a viable solution to the challenges identified in the literature to automating forest machinery. For example, when studying the requirements for an autonomous robot cleaning young stands, Vestlund and Hellström (2006) found that obstacle avoidance and identification of the stems of target trees were the greatest challenges to achieve autonomous operation. Similarly, Hellström et al. (2009) list the detection of obstacles as one of the most important challenges for automating forest machinery. They observe that in addition to the ability to find and detect objects blocking the machine's path, it is also necessary to consider *negative obstacles*, for instance a ditch or a steep slope.

The minimal requirements for a perception system for an autonomous forest machine involve the robot's ability to sense a) its own state and b) the surrounding forest. To adequately sense its own state, a forest machine must know its own position and orientation in the forest. Then, it can operate in the right forest plot given the location of the plot borders. It must also know its own inclination to avoid falling and crashing on sloped terrain. Furthermore, it must be aware of the position of its crane and tool and how they are oriented in order to effectively control them. The surrounding forest must be measured and modeled so that the robot knows where the forest floor is traversable and where to cut free space for strip roads. Naturally, it must also be capable of detecting, classifying, and

modeling the trees to be able to cut them, safely fell them, delimb them, and buck them to the right length. Finally, the logs must be piled on the forest floor, from where they are later carried to the roadside landing.

The role of the user interface is important in research aiming for semi-autonomous operation (e.g., Westerberg & Shiriaev, 2013; Westerberg, 2014). In semi-autonomous operation, information must flow seamlessly between the robot and the user. For example, augmented and virtual reality have been seen as plausible technologies to share the robots knowledge with the human operator (Westerberg & Shiriaev, 2013; Nordlie & Till, 2015; Palonen, 2016).

1.2 Research Objectives

The main target of this research is to increase the level of automation in forest machinery. In particular, it strives to develop sensors and methods to enable the forest machine to perceive its own state and the surrounding environment.

One major challenge is accurate and reliable measurement of the forestry crane boom tip and tool position while the machine is in operation. This problem can be divided into two challenges: 1) measurement of the posture of the forestry crane, i.e., the position and orientation of each of the flexible segments of the crane, and 2) measurement of the orientation of the tool which is hanging freely from the end of the crane. In addition, the position and orientation of the forest machine itself must also be measured to fix relative crane tip and tool positions to world coordinates. This can also be divided into two parts: A) the orientation of the body of the forest machine with respect to the Earth coordinate frame (i.e., the attitude) and B) the position and heading of the machine.

The surrounding environment must be observed to enable safe and productive autonomous or semi-autonomous work in it. The most important challenges to enabling autonomous operation are the reliable detection of all trees and the accurate classification of their species (Vestlund & Hellström, 2006). These are both challenging tasks, especially in dense younger forests where different tree species compete against each other. For autonomous operation, it is also important to be able to measure the forest floor and the free traversable space around the machine. This is a challenge in forests containing rocks, ditches, pits, large stumps and trees, where the automated machine should still be able to avoid any collision and navigate safely (see, e.g., Ahtiainen et al., 2017; Ruetz et al., 2022)[7].

As many of these challenges are extremely complex, full autonomy of forest machinery is difficult to achieve. Therefore, semi-autonomous so-

[7]Note that this dissertation does not study traversability nor object detection as such but instead sensors and perception methods to provide suitable data.

lutions, where the machine operator and the robot cooperate, are favored in the attempts to increase the productivity of forest machinery. In these semi-autonomous solutions, a new challenge emerges, however. The operator and robot must share the same knowledge about the problem at hand. This may be solved by visualizing data for operators to allow them to make better decisions. In addition, suitable selection of measurements about the machine and surrounding environment may enhance the operator's ability to perform various tasks.

To achieve shared knowledge between the operator and the robot, a user interface that is more efficient than traditional displays, touchscreens, joysticks, keyboard, and mouse is required. Virtual and augmented reality have been seen as plausible means of intuitively showing the machine operator information that the robot knows. (Westerberg & Shiriaev, 2013; Nordlie & Till, 2015; Palonen, 2016). Augmented reality, on the other hand, creates some new challenges, as it is difficult to measure the position and orientation of the operator's head in the forest machine cabin. Furthermore, a sufficiently large field of view (FoV), an adequately low time delay in head pose estimation, and sufficiently rapid visualization without excessive jitter on the head-mounted display (HMD) are necessary to avoid motion sickness (Lawson, 2015, pp. 567–568). Naturally, head movement and visual-field movement must correspond with each other.

To find solutions to all these challenges, the thesis is divided into 10 research questions. They fall into two categories: The first seven search for solutions to individual problems arising directly from these challenges (Q_1–Q_7) and the last three relate to more general machine perception solutions by combining the individual studies (Q_8–Q_{10}):

Q_1 Can an optical method, for example, a laser scanner, be used to advance the state of the art in estimating the posture of a forestry crane and the position of the boom tip?

Q_2 Can a low-cost inertial measurement unit (IMU) reliably measure the attitude of a forest machine in the presence of noise and non-gravitational accelerations?

Q_3 How can the rotator link mechanism be instrumented to measure the three-dimensional orientation of the tool of the forest machine when it is freely swaying from the end of the crane?

Q_4 Is it plausible to present sensory information in real-time from the perception system of the forest machine to the machine operator using augmented reality in the forest machine cabin?

Q_5 Is it possible for a forest machine to operate autonomously in an environment that presents significant challenges to the human operator?

Q_6 How can tree species be identified and classified in a young forest

using forest-machine-mounted sensors?

Q_7 How can a forest-machine-mounted laser scanner be used to measure the surrounding forest and segment trees and ground surface from the measurements?

Q_8 What kind of trade-offs help to enable real-time capable methods?

Q_9 What benefits are offered by different actuated laser scanning configurations?

Q_{10} How can sensor fusion with inertial measurements benefit machine perception for forest machinery?

Table 1.1. Relationship between the Research Questions (Q_1–Q_{10}) and the Publications (PI–PVII) summarized in Section 1.3

	PI	PII	PIII	PIV	PV	PVI	PVII
Q_1	x						
Q_2		x					
Q_3			x				
Q_4				x			
Q_5					x		
Q_6						x	
Q_7							x
Q_8	x	x	x	x	x		
Q_9	x					x	x
Q_{10}		x	x	x	x	x	x

The research questions and their relationship to the publications is shown in Table 1.1. The first seven questions (Q_1–Q_7) are covered one by one in publications I to VII. The last three questions (Q_8–Q_{10}), on the other hand, contain a wider scope than any of the individual publications. The publications are briefly summarized in the next section to provide an initial perspective on the research included in this thesis.

1.3 Summary of Publications and Contribution

Publication I (*Forestry crane posture estimation with a two-dimensional laser scanner*) proposes a novel method to estimate the posture of a flexible forestry crane by using a two-dimensional laser scanner, a rotary encoder, and two cylindrical metal targets attached to the crane boom. The method

uses a particle filter to track the two cylindrical targets in the laser scanner data with the help of digital control signals from the forestry crane, which are read from the vehicle's controller area network (CAN) bus. In contrast to conventional particle filter implementations, the proposed method uses an approximate inverse measurement model to track the two cylindrical targets in scanner measurements. This avoids the computational burden normally required in similar particle filter implementations of ray-casting each laser beam in each scan. This approximation enables the particle filter to be computed in real time with low-computational requirements (Q_8).

The proposed method is shown to work robustly in real time in the presence of obstructions and noise under practical operating conditions. Publication I advances the state of the art in estimating the posture of a forestry crane and the position of the boom tip (Q_1). As the method observes the crane tip position directly, the crane posture estimate therefore also observes the bending of the flexible crane. This is the first method to estimate the posture of a flexible forestry crane in real time using an optical sensor in such a way that it is robust with respect to line-of-sight obstructions. By using a vertical laser scanning configuration with the scanner attached on the side of the forestry crane boom, Publication I allows most of the scanner data to be simultaneously used to measure the environment around the working area in 3D in the same coordinates as the crane position (Q_9).

Publication II (*A DCM based attitude estimation algorithm for low-cost MEMS IMUs*) proposes an attitude estimation algorithm for fusing triaxial accelerometer and gyroscope measurements from an inexpensive micro-electro-mechanical system (MEMS) inertial measurement unit (IMU). The proposed adaptive, extended-Kalman-filter-based algorithm enables on-line gyroscope bias estimation with only triaxial acceleration and angular velocity measurements. The method also estimates the number of temporary (or transient) non-gravitational accelerations and adapts its attitude and bias estimation accordingly. In practice, this enables robust attitude estimation, for example, on a mobile phone or vehicle such as a robot or a forest machine, which are susceptible to accelerations, vibrations, and shocks. Furthermore, an opensource code and a test dataset are published for the proposed algorithm (Hyyti, 2015).

The contribution of Publication II is that it integrates many key modifications into the same extended Kalman filter, enabling attitude estimation in challenging conditions (Q_2). These modifications include: 1) adaptation to temporary or transient accelerations, 2) use of the lowest row of a direction cosine matrix (DCM, i.e., a rotation matrix) as an attitude estimate to exclude the otherwise problematic estimation of unknown rotation around the gravity vector (i.e., the heading), 3) use of gyroscope signals as control inputs in the filter, 4) allowance of variable time steps in the filter, and 5)

inclusion of online gyroscope bias estimation in the filter. The publication's contribution also lies in the fact that all this can be achieved using measurements from a low-cost integrated triaxial accelerometer and gyroscope MEMS chip alone. This method also contributes to publications III–VII, which require the orientation of the sensor (or at least part of the system) to be known (Q_{10}).

Publication III (*Sway estimation using inertial measurement units for cranes with a rotating tool*) proposes a novel method to estimate the pose of a freely swaying tool (e.g., a gripper or a cleaning tool) mounted on the tip of a forestry crane. The method uses a Kalman filter to fuse signals from two similar low-cost MEMS IMUs to estimate the three angles of a rotator link and a rotator motor: swaying angles sideways and back-and-forth, and the rotation angle of the hydraulic motor, called a rotator. The first IMU is mounted on the boom tip before the rotator link, and the second IMU is fitted to the tool hanging from the rotator. The computation occurs in real time on an embedded computer in the vicinity of the first IMU and the estimated joint angles and angular velocities are transmitted to the vehicle through a CAN bus. Therefore, only one cable is required to be fed through or pass the rotator link and the rotator.

The main contribution of Publication III is that its proposed method enables practical measurement of the tool orientation (Q_3). It is almost impossible to build traditional mechanical angle instruments into the rotator link mechanism, since any constructions must be mounted inside the steel structure to avoid damage when parts strike, for instance, the ground, logs, or trees. In addition, the sensors and cabling must be completely sealed against water and withstand freezing temperatures. Furthermore, the joint has three degrees of freedom and contains a hydraulic motor in an extremely compact space, which makes it challenging to integrate three angle position sensors inside the same structure. Therefore, these angles of the tool hanging from the tip of a forestry crane are seldom measured at all, which significantly limits computer-controlled or robotic use of forest machinery. The proposed method is the first known practical proposal to measure the orientation of the freely hanging and swaying tool of a forest machine.

Publication IV (*Augmented reality in forest machine cabin*) is the first published demonstration of a real-time augmented reality system in a forest machine cabin. Here, a head-mounted camera and an IMU are used to estimate head pose using a Kalman-filter-based sensor-fusion method. Laser scanner measurements, which are collected with the instrumentation of Publication I, are drawn on the video from the operator's helmet-mounted camera in real time. The demonstration aims to demonstrate that augmented reality is a viable user interface for a semi-autonomous forest machine, since it represents a feasible means of visualizing measurements and derived models intuitively for the human operator. The main contri-

bution of Publication IV is that it shows how augmented reality can be achieved by tracking the pose of the forest machine crane and the pose of the operator's head in a forest machine cabin (Q_4).

Publication V (*Real-time detection of young spruce using color and texture features on an autonomous forest machine*) demonstrates a real-time machine perception system for detecting young spruce trees among other vegetation using extracted color and texture features. Here, a machine vision camera is mounted in the center of a point cleaning tool to look downwards and observe the trees and the forest floor. The publication's main contribution is the use of a fully instrumented forest machine research prototype (Kalmari, Pihlajamäki, Hyyti, Luomaranta, & Visala, 2013) to perform the cleaning operation autonomously based on the real-time detection results (Q_5). Publication V documents the first demonstration where a perception system is used within a control loop in a forestry cleaning operation performed without any human intervention by an autonomous forest machine.

Publication VI (*Detection and species classification of young trees using machine perception for a semi-autonomous forest machine*) proposes a method for young tree detection and species classification using a machine perception system. The system combines a machine vision camera and a three-dimensional (3D) lidar built from an actuated laser scanner (Q_9). This is a novel approach to the automatic detection and classification of young trees in a forest. The method segments separate trees using lidar data before attempting to classify their species. The main contribution of the work is that it exemplifies how lidar and camera data may be combined and used for tree detection and classification purposes (Q_6).

Publication VII (*Feature based modeling and mapping of tree trunks and natural terrain using 3D laser scanner measurement system*) proposes methods for using a rotated two-dimensional (2D) laser scanner to collect 3D data from forests and to model and map the forest environment in real time. The method employs a 45° tilted rotating 2D laser scanner in a novel way to efficiently classify the collected point cloud into ground surface, tree edge, and center-of-tree-stem classes (Q_9). The method is efficient to compute in real time. As demonstrated in the study, these classes can later be efficiently used to search and segment individual trees from the point cloud. The publication's main contribution is its demonstration of how a forest-machine-mounted laser scanner can be used to measure the surrounding forest and segment trees and ground surface from the measurements (Q_7). This work also represents an example of how data from an actuated 2D laser scanner can be combined into a 3D point cloud around the vehicle.

1.4 Thesis Outline

After the introductory sections in Chapter 1, the contents of this thesis are divided as follows. Chapter 2 reviews machine perception in the forest environment. It discusses the probabilistic sensing and sensor fusion methods that can best utilize the noisy sensor data available in forest machinery. Two interconnected problems are identified: 1) sensing the robot/machine's own state or motion (i.e., proprioceptive perception) and 2) sensing the surrounding environment (i.e., exteroceptive perception). Both of these problems must be solved simultaneously, and the solutions often require measurements from multiple different sensors. Therefore, in addition to different perception modalities, probabilistic sensor fusion methods are important in this context.

Chapter 3 presents the proposed proprioceptive perception methods and their sensor setups. The novel sensor setups include minimal instrumentation for crane posture estimation and dual-IMU instrumentation of a forest machine tool. They are explained in more detail than in the published articles to facilitate the industrial adaptation of the methods developed in the studies. The proprioceptive perception methods include a particle filter for crane posture estimation in Publication I, and three extended Kalman filters, one for robust and adaptive attitude estimation using an IMU in Publication II, another for tool swaying angle estimation in Publication III, and the third for operator's head pose estimation in a forest machine cabin in Publication IV.

Chapter 4 presents the exteroceptive perception methods and their sensor setups proposed in this thesis for autonomous and semi-autonomous forest machinery. The sensor setups include three different actuated laser scanner configurations and the machine vision instrumentation for the point cleaning tool to detect young spruce trees. In the sensor setups, the practical hardware innovations and industrial applicability are elaborated. The exteroceptive perception methods include real-time machine vision for young spruce detection in Publication V, young tree detection and species classification in Publication VI, and tree stem and ground model estimation in Publication VII.

In Chapter 5, the benefits and limitations of the proposed perception system are discussed, the key findings are identified, the sensor and method selection is justified, and the research questions are answered. Finally, Chapter 6 concludes the work.

2. Machine Perception in Forest

In this chapter, the state of the art in machine perception relevant for autonomous forest machinery is reviewed. Most state-of-the-art methods for machine or robot perception are probabilistic, aiming to use all available information (see Machine Perception under Section 1.1 for details). Furthermore, many perception setups combine multiple similar or different sensors, so fusing signals and information from various sources is essential.

Here, we may learn something from the way nature has solved similar problems. Recent studies on perception in animals and humans indicate that it is a result of a probabilistic inference process (e.g., Ma, Beck, Latham, & Pouget, 2006; Doya, Ishii, Pouget, & Rao, 2007; Girshick, Landy, & Simoncelli, 2011). These studies have suggested that the brain somehow represents and manipulates uncertain information, which can be described in terms of probability distributions (J. F. Ferreira & Dias, 2013, p. 4).

Analogously to animals, the perceptual capabilities required in an autonomous forest machine can be divided into internal (i.e., *proprioceptive*) and external (i.e., *exteroceptive*) senses (Spero, 2004; Campbell et al., 2018; Yeong, Velasco-Hernandez, Barry, & Walsh, 2021). Proprioceptive senses, such as the vestibular sense (i.e., the sensations of body rotation and of gravitation and movement) help to balance the vehicle and measure its *attitude*, the relative orientation of the Earth and body frames. Additionally, the kinesthetic sense (i.e., the positions and movements of one's skeletal joints) also help the forest machine know where its crane and tool are located and how they may be moved. For example, this includes the posture of the crane and the pose of the tool. *Pose* comprises the position and the orientation of an object in 3D space and *posture* consists of the relative poses of all parts of a flexible structure, such as the hydraulic crane of a forest machine.

Exteroceptive senses are required to detect relevant objects, such as trees, rocks, and other obstacles, and to find safely passable ground and avoid sinking into mud. In autonomous vehicles, exteroceptive senses

are usually built using sensors, such as machine vision cameras, radars, sonars, or lidars (Spero, 2004; Campbell et al., 2018; Yeong et al., 2021). These sensors collect a large amount of data from the surrounding environment, but their observations contain significant uncertainties and noise. The environment and the sensors are usually modeled to make sense of the measurements. A sensor model is a mathematical description of the relation of sensor measurements to real objects in the environment. For example, if a measurement can be related to a specific location, the data may be represented as a *point cloud*, which is a high-resolution, densely spaced network of 3D points (Campbell et al., 2018). The point cloud can also contain some extra information, such as the color or intensity attached to each point. Alternatively, a *map* can be constructed from the surrounding environment to pinpoint the location of each detected object.

As explained earlier in the Background (Section 1.1), a forest is an uncontrolled, less structured environment, where sensors readings are uncertain and difficult to interpret. A perception system must cope with imperfect data, as forest machines operate in an environment where even the best remote sensing equipment struggles to measure trees precisely (e.g., X. Liang et al., 2018; J. Hyyppä et al., 2018) and where even state-of-the-art navigation sensors lack sufficient positioning accuracy (e.g., Kaartinen et al., 2015). When measurements tend to contain significant amount of noise and errors, it is beneficial to focus on probabilistic methods that can effectively combine multiple uncertain measurements to construct more reliable perception. When algorithms are built to cope with such errors and noise in the data, instead of focusing on the best possible sensors, lower cost but less accurate sensors can also be utilized for the task.

Cost is one of the limiting factors for adding sensors to commercial forest machinery. Therefore, expensive sensors (e.g., military grade IMUs, custom built radars, high quality 3D lidars, multi- or hyper-spectral cameras, and X-ray sensors) are excluded from consideration. Instead, low-cost sensors, such as MEMS IMUs, machine vision cameras, and laser scanners are the focus of this work. They are introduced in their own sections after the following section on probabilistic sensing and sensor fusion.

2.1 Probabilistic Sensing and Sensor Fusion

All measurements contain uncertainty. Uncertainty arises from sensor limitations, noise, and the fact that most interesting environments are complex and largely unpredictable (Thrun, 2000). For example, our forest machine uses a lidar sensor (see Lidar under Section 2.4) to measure ranges to nearby objects by timing a time-of-flight (ToF) delay between emitted and received laser pulses. If we aim the laser beam towards a tree, assume that we hit the tree, and wait for the returning echo, the range can

be computed using the ToF equation (J. Shan & Toth, 2018, p. 3),

$$r = \frac{vt}{2}, \tag{2.1}$$

where r is the range or slant distance, v is the speed of light in air, and t is the time interval measured between emitted and received pulses. Traditionally, the classical measurement error theory[8] concerns only random errors in the measurement, and results in a probabilistic model, the normal distribution around the most likely value (Rossi, 2014, p. 26). However, in the forest, we need something else instead. The distance measured varies considerably depending on whether the laser pulse has been reflected back from branches or needles at the front or rear of the tree, or whether the pulse has hit the trunk. The laser pulse might also collide with other objects in between the target and the sensor, which might shorten the range measurement drastically. Furthermore, it is possible that the laser pulse is reflected from an obstructing object in an unknown direction, resulting in the loss of the pulse or a random range measurement.

In the previous example, a simplistic deterministic algorithm would use a single range measurement and the equation (2.1) to resolve the distance to the tree. Because of the high number of possible error sources in the range measurement process, the range might vary significantly between measurement events. Therefore, it might be inadvisable to trust any single measurement. By contrast, a probabilistic algorithm is able to model uncertainty in the measurement process and combine multiple measurements. It computes a probability distribution over what might be the situation in the world instead of using an uncertain measurement per se (Thrun, 2000). This probabilistic computation is usually performed using a Bayesian framework, as explained in the following subsections.

Detection of an Event

According to the Bayesian worldview, probability is seen as the measure of belief or confidence in the occurrence of an event (Davidson-Pilon, 2015, p. 2). In the Bayesian context, an event is whatever can be precisely described by a proposition (D'Agostini, 2003, p. vii). The event can also be defined as a set of outcomes (Forsyth, 2017, p. 55). If it contains just two possible outcomes (e.g., true or false), it is called as a binary event. For instance, whether or not the input data contain a tree is an illustrative example of a binary event. The same example is followed throughout the chapter.

Even before measurement occurs, a Bayesian observer assigns a belief to the event based on their previous knowledge of the world. If they assign a belief of 0 to an event, they believe with absolute certainty that the

[8] Developed mainly thanks to the contributions of Carl Friedrich Gauss (1809, 1823) and Pierre-Simon Marquis de Laplace (1812).

event (e.g., tree detection) will not occur, and, conversely, assigning a belief of 1 implies absolute certainty that the event will occur. Naturally, the belief may be any rational number between these two extremes. To align with traditional probability notation, the probability of an event A can be expressed as $p(A) \in [0,1]$. A belief based on previous knowledge is called a *prior probability* (Davidson-Pilon, 2015, p. 2).

When our Bayesian observer obtains new information (for example, evidence through a set of range measurements), they now know more about the occurrence of the event and can change their belief. The updated belief can be denoted as $p(A \mid Y)$, which is interpreted as the probability of A given (conditioned on) the evidence Y (Rossi, 2014, p. 95). This is called the posterior probability (Davidson-Pilon, 2015, p. 3). The belief is updated from prior to posterior probability via the following equation, known as Bayes' rule (Davidson-Pilon, 2015, p. 5; Bar-Shalom, Li, & Kirubarajan, 2001, pp. 47–48):

$$p(A \mid Y) = \frac{p(Y \mid A)p(A)}{p(Y)}, \tag{2.2}$$

where $p(A)$ is the prior probability of the event, $p(Y \mid A)$ is the sampling distribution (i.e., the probability of evidence when the event A occurs), and $p(Y)$ is the prior probability of the evidence. After this step in the tree detection example, the observer now possesses an updated belief as to whether a tree is present. This belief combines prior knowledge with information acquired through an uncertain measurement.

The term *sampling distribution* is used for $p(Y \mid A)$ in Equation (2.2), when the event A is fixed. Instead, the same term may be considered a fixed measurement in the light of different hypotheses of events $\{A, A', A'', \ldots\}$. In its dependence on A for fixed Y, $P(Y \mid A)$ is called a *likelihood* (Jaynes, 2003, p. 89). The likelihood of event A (denoted as $\mathscr{L}(A)$) is not itself the probability of A. Instead, it is a dimensionless numerical function which, when multiplied by a prior probability and divided by a normalization factor, may become the probability of A. If these are constant, they are often omitted, and the proportionality between the posterior probability and the likelihood remains,

$$p(A \mid Y) \propto \mathscr{L}(A). \tag{2.3}$$

For example, in the tree detection case, the likelihood of tree detection could be computed for our input data. It would tell us how likely our different hypotheses of a tree are, given the measurements.

If the Bayesian observer possesses multiple different measurements, assuming that the measurements are conditionally independent (Thrun et al., 2005, p. 15), the observer can sequentially combine them all by just repeating Equation (2.2) for each piece of evidence Y using the latest posterior probability as a prior for the second measurement update, and so forth. If the observer possess knowledge about all probabilities, they can optimally integrate the available information.

Continuous Variables and Sensor Measurements

Contrary to the example of deciding on the presence or absence of a tree, measurements are rarely binary events. Rather, they are commonly continuous variables (e.g., readings from a measuring tape) or at least discrete variables with such fine discrete steps that the value can be treated as being drawn from a continuous probability distribution (e.g., a range measurement from a laser scanner). In the Bayesian framework, we usually make an assumption that these measurements are continuous random variables, i.e., samples from a continuous random process that has a probability density function (PDF) (Thrun et al., 2005, p. 10; Rossi, 2014, p. 105). This allows the Bayesian rule explained above (2.2) to be also used for continuous variables.

Electrical devices that automatically and repeatedly take measurements from the robot and its environment are called *sensors*. They convert the physical and chemical variables of a process or an installation into electrical (at first, usually analog) signals (Placko, 2007, p. 41). As a basic principle, the sensor converts measurand m into an electrical variable called s. The relation that joins these is a function $s = F(m)$, which may be linear or nonlinear depending on the physical law determining the sensor, the structure of the sensor, and the sensor's environment (Placko, 2007, p. 42). Sensor-measured analog signals can then be transformed into a digital form using an analog-to-digital converter. Several such digital variables can be combined as a single measurement. For example, laser scanners combine multiple sensor readings taken at different scanning angles as a 2D scan. Similarly, cameras sample received light intensities on each pixel, and together all pixel values form an image.

In a probabilistic framework, noise and inaccuracies in sensor measurements are explicitly modeled. These models account for the inherent uncertainty in the robot's sensors (Thrun et al., 2005, p. 121). The inability of the sensor to provide exact values for the assumed measurand is often paraphrased as sensor noise. These errors can be caused by random electrical signals or inaccuracies in the electronics and computation. In a probabilistic framework, possible errors in the measurement process can also be included in the sensor model. In a forest environment, these could be, for example, a false assumption of the target the laser beam has hit. As explained at the beginning of this chapter, such an error could provide overly short range measurements if the laser pulse has struck something else before the assumed target. Alternatively, the range can also be far larger than expected if the laser beam has missed the tree and, instead, hit something else behind it.

In a probabilistic framework, in addition to sensors, the environment in which a measurement is generated must also be specified. A common way to model the world is to use a *map* representation. A map of the

environment is a list of objects, indexed in one of two ways, namely *feature-based* and *location-based* (Thrun et al., 2005, p. 123). A feature-based map is usually represented as a list of detected objects (such as trees or rocks) or features (such as coded representations of sample points), and a location-based map is usually represented as a grid of coordinates (e.g., a voxel grid) that list the occupancy of all locations in the world. The grid can have one or more dimensions. Usually, 2D or 3D grids are used.

The internal representation of the map can be *metric* or *topological* (Thrun, 1998). In the more common metric framework, objects are placed with precise coordinates. This representation is very useful in robotics, since the map is easier to build, represent, and maintain, it allows geometric place recognition, it is view-point independent, and facilitates computation of shortest paths on the map. The downside is that metric approaches are more space-consuming, computationally inefficient, and they require precise location and orientation of the robot to be known.

The metric map may be constructed recursively by concatenating measured objects or modifying the occupancy grid over time, given that the robot's location and orientation are precisely known. If the pose is not known, which is the usual case, it must also be ascertained in the process. Then the mapping problem becomes a significantly more challenging "chicken-and-egg" problem, since the robot's own pose also affects the location of the sensed objects on the map. This problem is termed simultaneous localization and mapping (SLAM) (Thrun et al., 2005, p. 222). SLAM is a much-researched topic (e.g., see the following reviews, Spero & Jarvis, 2007; Lu, Hu, & Uchimura, 2009; Khairuddin, Talib, & Haron, 2015; Taheri & Xia, 2021). It is explained in more detail at the end of Section 2.2.

The topological framework, on the other hand, only considers places and relations between them. Often, the distances between places (or objects) are stored. The map can be represented as a graph, in which the nodes corresponds to places and arcs correspond to the paths between them. This approach is difficult to construct and maintain in large-scale environments if sensor information is ambiguous (Thrun, 1998). Furthermore, the recognition of places and objects is often difficult and sensitive to the point of view. However, precise positioning of the robot is not needed while constructing a topological map.

State Estimation

In control theory, a system can be modeled mathematically using state variables to describe the *state* of that system. The states of a system are those variables that provide a complete representation of its internal condition or status at a given instance in time (Simon, 2006, p. xxi). If the states develop with respect to time, the system is dynamic. A controlled

dynamic system may be described with equations,

$$\begin{aligned} \mathbf{x}_k &= f_{k-1}(\mathbf{x}_{k-1}, \mathbf{u}_{k-1}, \boldsymbol{\nu}_{k-1}) \\ \mathbf{y}_k &= h_k(\mathbf{x}_k, \boldsymbol{\upsilon}_k), \end{aligned} \tag{2.4}$$

where k is a time index (and $k-1$ is the previous one), $\mathbf{x}_k$ is the state, $\mathbf{u}_k$ is a known control input, $\boldsymbol{\nu}_k$ is the process noise, $\mathbf{y}_k$ is the measurement, $\boldsymbol{\upsilon}_k$ is the measurement noise, and functions $f_k(\cdot)$ and $h_k(\cdot)$ are time-varying nonlinear system and measurement equations, respectively (Simon, 2006, p. 463; Bar-Shalom et al., 2001, p. 372).

For example, the attitude of a forest machine may be modeled with a rotation matrix (which can be constructed through a few different formalisms described later in Section 2.2). At least three state variables are required to represent rotation in 3D space (e.g., Euler angles) (Diebel, 2006). Here, the function $f_k(\cdot)$ in Equation (2.4) would predict how the attitude represented by the state variables changes from the current time step to the next. The function $h_k(\cdot)$, on the other hand, would model the measurement process for this system given the state and uncertainties.

A state is deemed complete if it is the best predictor of the future (Thrun et al., 2005, p. 17). This means that knowledge of past states, measurements, or given control inputs carry no additional information to predict the future more accurately. Such systems are commonly known as Markov processes and, in discrete systems, Markov chains[9] (Bharucha-Reid, 2012, p. 4). However, in practice, it is impossible to specify a complete state for any realistic system (Thrun et al., 2005, p. 18). A complete state includes all aspects of the environment and the robot that may exert any impact in the future. Therefore, all practical implementations are limited to a small subset of all state variables, i.e., an incomplete state. For example, the previous attitude-estimation example assumed that the body of the forest machine is rigid, which means that it does not bend or contain any moving parts. Otherwise, more parameters would be required to explicitly determine the posture of a non-rigid object. These parameters might be difficult to define as demonstrated, for instance, in Publication I. Therefore, even for this limited attitude estimation example, the complete state is too difficult to obtain, and instead an approximated incomplete model must be used.

The process of inferring the best value of a quantity of interest from indirect, inaccurate, and uncertain observations is called *estimation* (Bar-Shalom et al., 2001, p. 1). For a mobile robot, a large amount of important knowledge about itself and the environment is not directly measurable. Instead, the robot must rely on its sensors to gather the relevant information. As noted earlier, sensors only collect partial information about the

[9] Named after the Russian mathematician A. A. Markov (1856–1922), who introduced the concept of chain dependence (Bharucha-Reid, 2012, p. 4).

required quantities, and their measurements are corrupted by noise. State estimation aims to recover these (partially) hidden state variables from the data (Thrun et al., 2005, p. 9). For example, the robot is unable to directly measure its attitude in the world. Instead, accelerometers and gyroscopes are usually used to estimate the selected state variables representing the attitude, as employed in Publication II.

In the estimation process, state estimation algorithms are used to compute belief distributions over all states $\mathbf{S}$ (Thrun et al., 2005, p. 9). A belief reflects the robot's knowledge about the state of the world (Thrun et al., 2005, p. 22). At time index k, the *belief distribution* can be represented with a conditional probability distribution, $p(\mathbf{x}_k \mid \mathbf{Y}_k, \mathbf{U}_{k-1})$, which is a PDF of the state $\mathbf{x}_k$ conditioned on all measurements in a time series $\mathbf{y}_1$, $\mathbf{y}_2$, ..., $\mathbf{y}_k$ and all given control inputs $\mathbf{u}_1$, $\mathbf{u}_2$, ..., $\mathbf{u}_{k-1}$ (Simon, 2006, p. 463; Thrun et al., 2005, p. 23). Assuming that the process is Markovian[9], the current state values contain all knowledge in the system (Bar-Shalom et al., 2001, p. 195). Then only the current round measurements and last control inputs are required. For these systems, the belief can be simplified as $p(\mathbf{x}_k \mid \mathbf{y}_k, \mathbf{u}_{k-1})$.

The belief is usually updated in two distinct phases, which are termed a *prediction*[10] and a *measurement update* (Thrun et al., 2005, p. 23). In the prediction step, the state of the system is updated using a prior belief of the state, a model of the system, and an optional control signal, which may have been given to the robot. In the measurement update step, sensor signals and a sensor model are used to update the belief. These steps are applied in turn to update the belief for each time instance in a series.

Bayes Filter

The belief update can be formulated as two distinct equations forming the most general algorithm for calculating beliefs, called the Bayes filter algorithm (Thrun et al., 2005, pp. 23–24; Simon, 2006, p. 465). The prediction step combines the last round belief $p(\mathbf{x}_{k-1} \mid \mathbf{y}_{k-1}, \mathbf{u}_{k-2})$ and the model to predict state values with the optional known control inputs, $p(\mathbf{x}_k \mid \mathbf{x}_{k-1}, \mathbf{u}_{k-1})$. The integration results in an a priori belief,

$$p(\mathbf{x}_k \mid \mathbf{y}_{k-1}, \mathbf{u}_{k-1}) = \int p(\mathbf{x}_k \mid \mathbf{x}_{k-1}, \mathbf{u}_{k-1}) p(\mathbf{x}_{k-1} \mid \mathbf{y}_{k-1}, \mathbf{u}_{k-2}) \, \mathrm{d}\mathbf{x}_{k-1}. \tag{2.5}$$

The measurement step computes the a posteriori belief using Bayes' rule in Equation (2.2) for the sampling distribution $p(\mathbf{y}_k \mid \mathbf{x}_k)$ and the a priori belief:

$$p(\mathbf{x}_k \mid \mathbf{y}_k, \mathbf{u}_{k-1}) = \frac{p(\mathbf{y}_k \mid \mathbf{x}_k) p(\mathbf{x}_k \mid \mathbf{y}_{k-1}, \mathbf{u}_{k-1})}{\int p(\mathbf{y}_k \mid \mathbf{x}_k) p(\mathbf{x}_k \mid \mathbf{y}_{k-1}, \mathbf{u}_{k-1}) \, \mathrm{d}\mathbf{x}_k}. \tag{2.6}$$

[10] Names such as time update (Bar-Shalom et al., 2001, p. 208; Simon, 2006, p. 127) or control update (Thrun et al., 2005, p. 23) are also used.

Note that the prior probability of the measurement (the denominator in Equation (2.6)) is computed through integrating over all possible state values. It is equivalent to normalization of the resulting PDF with a normalization constant η, and thus the equation may be simplified as (Thrun et al., 2005, pp. 23–24):

$$p(\mathbf{x}_k \mid \mathbf{y}_k, \mathbf{u}_{k-1}) = \eta p(\mathbf{y}_k \mid \mathbf{x}_k) p(\mathbf{x}_k \mid \mathbf{y}_{k-1}, \mathbf{u}_{k-1}). \tag{2.7}$$

In the previous equations, the control $\mathbf{u}$ is thought to occur after the update of the belief, so the newest control is always one step behind and has time index $k-1$.

Unfortunately, this general Bayes filter can only be implemented for very simple estimation problems. This is because we must either be able to perform the integration in Equation (2.5) and the multiplication in Equation (2.7) in closed form or restrict ourselves to finite state spaces, such that the integral in Equation (2.5) becomes a finite sum (Thrun et al., 2005, p. 24). Analytical, closed form solutions to Bayes filter equations are available for only a few special cases. In particular, if $f(\cdot)$ and $h(\cdot)$ are linear, and the initial state x_0, and noises $\{\boldsymbol{v}_k\}$ and $\{\boldsymbol{\upsilon}_k\}$ in Equation (2.4) are additive, independent, and Gaussian, then the closed form solution to Bayes filter equations is a Kalman filter (Simon, 2006, p. 465).

Kalman Filtering

Nonlinear and modified versions of the original Kalman filter (KF) are used in publications II, III, and IV. To introduce the basics of these algorithms, the original KF is quickly reviewed.

The KF[11] is probably the most studied technique for implementing the Bayes filter (Thrun et al., 2005, p. 34). It is used for filtering and predicting dynamic linear systems with additive, independent, and Gaussian noises and initial state. As an analytical Kalman filter is simpler to implement and significantly faster to compute than the Bayes filter or other numerical alternatives, it has become an extremely popular estimation approach despite the fact that the required assumptions are not always fully met. Briefly, if noises are zero-mean, uncorrelated, and white, then the KF is the best linear solution, and even if the noise is not Gaussian, the KF is still the optimal linear filter (Simon, 2006, p. 130).

Originally, the KF was derived for continuous time systems, but the discrete version is currently more common, since it is easier to implement on computers which operate in discrete time steps. In the discrete-time KF, the dynamic system is given by the following equations, which determine how the system changes from the previous time step $(k-1)$ to the current

[11] The Kalman filter was invented in the 1950s by Rudolph Emil Kalman (1960).

step (k) (Simon, 2006, p. 128):

$$\begin{aligned}\mathbf{x}_k &= \mathbf{F}_{k-1}\mathbf{x}_{k-1} + \mathbf{G}_{k-1}\mathbf{u}_{k-1} + \boldsymbol{\nu}_{k-1} \\ \mathbf{y}_k &= \mathbf{H}_k\mathbf{x}_k + \boldsymbol{\upsilon}_k,\end{aligned} \tag{2.8}$$

where $\mathbf{Q}_k$ is a state-prediction or time-update covariance of a noise process $\{\boldsymbol{\nu}_k\}$, and $\mathbf{R}_k$ is a measurement covariance of a noise process $\{\boldsymbol{\upsilon}_k\}$. The noises $\boldsymbol{\nu}_k$ and $\boldsymbol{\upsilon}_k$ are assumed to be white, zero mean, and uncorrelated. In Equation (2.8), $\mathbf{F}_k$ and $\mathbf{G}_k$ are matrices representing linear dynamic system models for the state $\mathbf{x}_k$ and control input $\mathbf{u}_k$, respectively. Similarly, $\mathbf{H}_k$ is a linear measurement model giving the measurement $\mathbf{y}_k$.

In the KF, the belief is represented with Gaussian distributions, i.e., the mean $\hat{\mathbf{x}}_k$ and the covariance $\mathbf{P}_k$, and the filter is initialized at the beginning ($k = 0$) as follows:

$$\begin{aligned}\hat{\mathbf{x}}_0 &= E[\mathbf{x}_0] \\ \mathbf{P}_0 &= E[(\mathbf{x}_0 - \hat{\mathbf{x}}_0)(\mathbf{x}_0 - \hat{\mathbf{x}}_0)^\mathsf{T}],\end{aligned} \tag{2.9}$$

where $E[\cdot]$ is the expected value operator (Simon, 2006, p. 128).

The Kalman filter is provided by the following equations, which are computed iteratively for each time step $k = 1, 2, \ldots$ (Simon, 2006, pp. 128–129). The first set of equations performs the prediction phase,

$$\hat{\mathbf{x}}_k^- = \mathbf{F}_{k-1}\hat{\mathbf{x}}_{k-1} + \mathbf{G}_{k-1}\mathbf{u}_{k-1} \tag{2.10}$$

$$\mathbf{P}_k^- = \mathbf{F}_{k-1}\mathbf{P}_{k-1}\mathbf{F}_{k-1}^\mathsf{T} + \mathbf{Q}_{k-1}, \tag{2.11}$$

where the estimated mean and covariance are predicted with the linear models $\mathbf{F}_k$ and $\mathbf{G}_k$. The minus sign as a superscript ($\cdot^-$) indicates the predicted estimate. For the measurement update, the so-called Kalman gain must be defined (Simon, 2006, p. 128):

$$\mathbf{K}_k = \mathbf{P}_k^-\mathbf{H}_k^\mathsf{T}\left(\mathbf{H}_k\mathbf{P}_k^-\mathbf{H}_k^\mathsf{T} + \mathbf{R}_k\right)^{-1}. \tag{2.12}$$

The a posteriori belief is then computed by incorporating the measurement $\mathbf{y}_k$ and the Kalman gain $\mathbf{K}_k$ with the measurement model $\mathbf{H}_k$ and measurement covariance $\mathbf{R}_k$:

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_k^- + \mathbf{K}_k\left(\mathbf{y}_k - \mathbf{H}_k\hat{\mathbf{x}}_k^-\right) \tag{2.13}$$

$$\mathbf{P}_k = (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\mathbf{P}_k^-(\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)^\mathsf{T} + \mathbf{K}_k\mathbf{R}_k\mathbf{K}_k^\mathsf{T}, \tag{2.14}$$

where (2.14) is the so-called Joseph stabilized version of the covariance update equation (Simon, 2006, p. 129). It was formulated by Peter Joseph in the 1960s, and it guarantees that $\mathbf{P}_k$ will always be symmetric positive definite, as long as $\mathbf{P}_k^-$ is symmetric positive definite. There also exists a computationally simpler version for updating the covariance, $\mathbf{P}_k = (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\mathbf{P}_k^-$, but it does not guarantee symmetry or positive definiteness for $\mathbf{P}_k$ (Simon,

2006, p. 129). If, for any reason (e.g., inaccuracies in the floating-point computation) the covariance becomes asymmetric or indefinite, the Kalman filter estimate might diverge quickly.

Kalman filter is attractive to use, since if the assumptions are met, it is the one estimator that results in the smallest possible standard deviation of the estimation error (Simon, 2006, p. 336). Therefore, it is called an optimal linear state estimator.

Most state estimation methods have been developed for linear dynamic systems (Kalman filter, Information filter, H_∞ filter, etc.), from which many are restricted to normally distributed belief distributions (Simon, 2006). Unfortunately, the world is rarely linear, and many distributions are non-Gaussian. Therefore, nonlinear methods are often required in many practical cases, such as when estimating the attitude (e.g., in Publication II) or a pose of a robot or its manipulator (e.g., in publications I and III).

Nonlinear Filtering

Nonlinear filtering is significantly more complex and difficult than linear filtering (Simon, 2006, p. 396). However, we are frequently required to use nonlinear techniques in cases where linear or Gaussian assumptions can not be made. Thus, some nonlinear estimation methods have become more widespread. These techniques include nonlinear extensions of the Kalman filter, unscented filtering, and particle filtering, from which nonlinear Kalman filtering is the most widespread approach to state estimation for nonlinear systems (Simon, 2006, p. 425).

The extended Kalman filter (EKF) is the most widely used nonlinear state estimation technique (Simon, 2006, p. 396). In the EKF, a nonlinear system is simply linearized around the Kalman filter estimate using a first-order Taylor series expansion[12] (Simon, 2006, p. 400). When the estimate changes, the nonlinear system is linearized around the updated estimate.

In a discrete-time EKF, the system model and the measurement model are nonlinear functions $f_k(\cdot)$ and $h_k(\cdot)$, respectively (Simon, 2006, p. 409):

$$\begin{aligned} \mathbf{x}_k &= f_{k-1}(\mathbf{x}_{k-1}, \mathbf{u}_{k-1}, \boldsymbol{\nu}_{k-1}) \\ \mathbf{y}_k &= h_k(\mathbf{x}_k, \boldsymbol{\upsilon}_k), \end{aligned} \tag{2.15}$$

where $\mathbf{x}_k$ is the state, $\mathbf{u}_k$ is the control input, $\mathbf{y}_k$ is the measurement, and $\boldsymbol{\nu}_k$ and $\boldsymbol{\upsilon}_k$ are the noises at time index k, similar to Equation (2.8). The filter is initialized in a similar manner to the linear version in Equation (2.9), but the prediction step in Equation (2.10) differs. In the EKF, the nonlinear model f_{k-1} is linearized around the previous estimate $\hat{\mathbf{x}}_{k-1}$ (Simon, 2006,

[12] The idea was originally proposed by a NASA engineer, Dr. Stanley Schmidt, to allow the KF to be applied to nonlinear navigation problems in space (Schmidt, 1962, 1966).

p. 409):

$$\mathbf{F}_{k-1} = \left.\frac{\partial f_{k-1}}{\partial \mathbf{x}}\right|_{\hat{\mathbf{x}}_{k-1}} \qquad \mathbf{L}_{k-1} = \left.\frac{\partial f_{k-1}}{\partial \boldsymbol{v}}\right|_{\hat{\mathbf{x}}_{k-1}}, \tag{2.16}$$

where $\mathbf{F}_k$ is the linearized model equivalent to the one in Equation (2.10), but $\mathbf{L}_k$ is a new term which models the linearization of the process noise at time index k.

The nonlinear model is used for state prediction without noise, and the linearized models are used for covariance prediction (Simon, 2006, p. 409):

$$\hat{\mathbf{x}}_k^- = f_{k-1}(\hat{\mathbf{x}}_{k-1}, \mathbf{u}_{k-1}, 0) \tag{2.17}$$

$$\mathbf{P}_k^- = \mathbf{F}_{k-1}\mathbf{P}_{k-1}\mathbf{F}_{k-1}^\top + \mathbf{L}_{k-1}\mathbf{Q}_{k-1}\mathbf{L}_{k-1}^\top, \tag{2.18}$$

In many cases the noise is assumed to be linear, and then $\mathbf{L}_k$ becomes an identity matrix and can be omitted from the equations. Thus, the covariance update may be simplified to the original equation in (2.11).

Similar to the prediction, the measurement update with the nonlinear measurement model $h_k(\cdot)$ must also be linearized. Linearization is performed around the predicted estimate (Simon, 2006, p. 409):

$$\mathbf{H}_k = \left.\frac{\partial h_k}{\partial \mathbf{x}}\right|_{\hat{\mathbf{x}}_k^-} \qquad \mathbf{M}_k = \left.\frac{\partial h_k}{\partial \boldsymbol{v}}\right|_{\hat{\mathbf{x}}_k^-}. \tag{2.19}$$

With the linearized models $\mathbf{H}_k$ and $\mathbf{M}_k$, the measurement update is performed in a similar manner to the linear Kalman filter in equations (2.12) and (2.13). However, instead of using the measurement covariance, the noise linearization with $\mathbf{M}_k$ is taken into account in a similar manner to Equation (2.17) with $\mathbf{L}_k$. Furthermore, if the measurement noise is also assumed to be linear, $\mathbf{M}_k$ can be assumed to be an identity matrix. Then the required addition of $\mathbf{M}_k\mathbf{R}_k\mathbf{M}_k^\top$ to the covariance update may be omitted and the original $\mathbf{R}_k$ used instead, as in equations (2.12) and (2.14).

The EKF enables the KF to be used for nonlinear systems, but it is still restricted by linearization and the assumptions of additive, independent, and Gaussian initial state and noises (Bar-Shalom et al., 2001, p. 394). The equations (2.16) to (2.19), previously shown above, assume that a linearized transformation of means and covariances is approximately equal to the true nonlinear transformation (Simon, 2006, p. 441). Therefore, the EKF may be used for slightly nonlinear problems if the initial error and noises are sufficiently small (Bar-Shalom et al., 2001, p. 387). However, when nonlinearity or errors increase, other means should be used instead.

There are many modifications of the EKF that aim to handle nonlinearity more accurately. One of these is a second order EKF, which also uses the second order term of Taylor series in its approximation of nonlinear distributions (see, e.g., Simon, 2006, pp. 419–420). Alternatively, in an iterated EKF, the measurement equation is iterated several times to obtain a more accurate point of linearization (see, e.g., Bar-Shalom et al., 2001,

pp. 404–406). Compared with attempting to linearize the whole distribution at once with a Taylor approximation, a single point is far easier to put through a nonlinear transformation, as performed in an unscented Kalman filter (UKF). Furthermore, it is straightforward to select a set of single points such that these points sample the same distribution. In an UKF, $2N$ sigma points are selected around the mean to sample the mean and distribution, where the N is the length of the state vector (see more details, e.g., in Simon, 2006, pp. 433–459).

When the problem is significantly nonlinear or distributions are significantly non-Gaussian (e.g., a distribution with two or more modes), KF-based filters should no longer be used. Then we must seek other means of using the Bayes filter introduced above (see equations (2.5) to (2.7)), which can handle nonlinearities and any form of noise.

An implicit solution to approximate the Bayes filter is to decompose the state space into a finite number of regions and represent the cumulative posterior for each region with a single probability value. For discrete state spaces, this is called a discrete Bayes filter and, for continuous state spaces, a histogram filter (Thrun et al., 2005, p. 68). In histogram filters, continuous space is discretized into finite regions, usually called histogram bins, and then the same equations as in the discrete Bayes filter may be applied. However, a histogram filter becomes computationally infeasible when the problem dimensions increase, since its computational complexity increases to $\mathcal{O}(N_c^2)$, where N_c stands for the number of cells in the filter (Blanco Claraco & Fernández-Madrigal, 2012, p. 238). This implies that if each dimension is split to N_b bins, in a three-dimensional problem, the computational cost already increases to $\mathcal{O}(N_b^6)$. This limits the use of the histogram filter to be used only in low-dimensional problems.

One solution to avoid this exponential rise in computational cost is to approximate the belief distribution only around selected interest points. This may be achieved, for example, with a particle filter (PF), which is the method applied in Publication I to estimate crane posture. In a PF, the posterior belief is approximated with a set of random state samples called *particles*, which are drawn randomly from the belief posterior (Thrun et al., 2005, p. 77). Similar to previously explained filters, a PF estimates the states of a hidden Markov model in a sequential fashion (Arulampalam, Maskell, Gordon, & Clapp, 2002; Candy, 2007). After each observation, the (hidden) state of the system is sequentially estimated from the posterior density of the state variables.

In a PF, the state can be denoted as a set $\mathbf{X}(k)$ at time index k, which contains N_p particles, $\mathbf{x}_i(k)$, where $i \in \{1,...,N_p\}$. All of these particles have an associated weight $w_i(k)$, which measures how well each particle fits to the estimated posterior PDF, i.e., their importance. In the original Sequential Importance Sampling (SIS) type particle filter, the weights are selected such that, together, the locations of the particles $\mathbf{x}_i$ and their

associated weights w_i best estimate the belief posterior of the current state and all previous measurements (Arulampalam et al., 2002; Candy, 2007).

Similar to the histogram filter, the PF is a statistical, brute-force approach that often works well for problems that are too difficult for the conventional Kalman filter (i.e., systems that are highly nonlinear or that contain multimodal distributions) (Simon, 2006, p. 461). They differ in the way these parameters are generated and the way they populate the state space. However, in cases where the PDF is concentrated on a small subset of the whole state space, even a small number of particles may sufficiently approximate the true PDF. However, in spite of being a brute-force approach, computing a small number of particles is significantly faster than computing the values in all histogram cells.

One of the simplest versions of the PF is a Sampling Importance Resampling (SIR) type filter[13], an approach first proposed by Gordon et al. (1993). In the SIR filter, a transition prior is used as the importance density in the weight update stage (Gordon et al., 1993; Arulampalam et al., 2002; Candy, 2007). This leads to a major simplification of the PF, as only the likelihood of the last measurement $\mathbf{y}(k)$, $\mathscr{L}(\mathbf{y}(k)|\mathbf{x}_i(k))$, is used to update weights $w_i(k) \in \mathbb{R}$ for each particle $\mathbf{x}_i(k)$ $(i = 1,\ldots,N_p)$:

$$w_i(k) = w_i(k-1)\ \mathscr{L}(\mathbf{y}(k)|\mathbf{x}_i(k)). \tag{2.20}$$

To use the SIR algorithm, all N_p particles must be initialized by drawing them from a known initial distribution $p(x_0)$ (Simon, 2006, p. 468). Then, for each time index, $k = 1,2,\ldots$, four steps are iterated sequentially:

a) A-priori particles are predicted using the known process equation f_k in (2.4), and the known PDF of the process noise $\boldsymbol{\nu}_k$,

b) Weight w_i is updated for each particle $\mathbf{x}_i$ using Equation (2.20) with the measurement model h_k and the measurement noise $\boldsymbol{\upsilon}_k$ from the state space model in Equation (2.4),

c) weights are normalized so that $\sum_{i=1}^{N_p} w_i(k) = 1$, and

d) if needed, the particles are resampled by generating a set of new particles with probabilities according to their normalized weights.

The benefit of the SIR method is that the importance weights are simple to evaluate, and the importance density can be easily sampled (Arulampalam et al., 2002). The disadvantage of the SIR method is that since neither earlier nor current round measurements are taken into account at the prediction phase, the particles are depleted much faster than in the original Sequential Importance Sampling (SIS) algorithm (Arulampalam et al., 2002). This means that a few particles will eventually account for a

[13]This PF type is also called a Bootstrap Particle Filter (BPF) (Gordon, Salmond, & Smith, 1993; Candy, 2007), as the key update stage of the algorithm (Bayes rule) is implemented as a weighted bootstrap (Smith & Gelfand, 1992).

significantly large share of the total weight. This is called a degeneracy problem. To avoid it, a high concentration of probability mass at a few particles should be avoided by, for instance, resampling (Gustafsson et al., 2002; Arulampalam et al., 2002).

Although resampling effectively deals with the degeneracy problem, it introduces a new problem, called the sample impoverishment problem. The diversity of the particles will tend to decrease after each resampling step because particles with large weights are likely to be drawn several times during resampling, while particles with minor weights are unlikely to be drawn at all. This means that the resampling process will select only a few (or even one) distinct a priori particles to become posteriori particles (Simon, 2006, p. 470). This problem is severe in the case of low process noise, as probability mass is drawn into a single point (Arulampalam et al., 2002). However, roughening, i.e., adding artificial process noise to the prediction step, tends to spread the PDF, which mitigates the effect (Simon, 2006, p. 470). In Publication I, the process noise is large, and therefore this simpler SIR type filter can be effectively used.

In summary, in a nonlinear system, (nonlinear versions of) the Kalman filter can be used for state estimation, but the larger computational effort required by the particle filter may provide better results. On the other hand, in a linear system that contains non-Gaussian noise, the Kalman filter is the optimal linear filter, but again the particle filter may perform better (Simon, 2006, p. 480). The unscented Kalman filter balances between the low computational effort of the Kalman filter and the high performance of the particle filter. As depicted in Figure 2.1, going from an (E)KF through a UKF to a PF increases the computational effort of state estimation. However, this only increases the accuracy for nonlinear or non-Gaussian systems (Figure 2.1A). For linear and Gaussian systems, the KF is the optimal choice (Figure 2.1B).

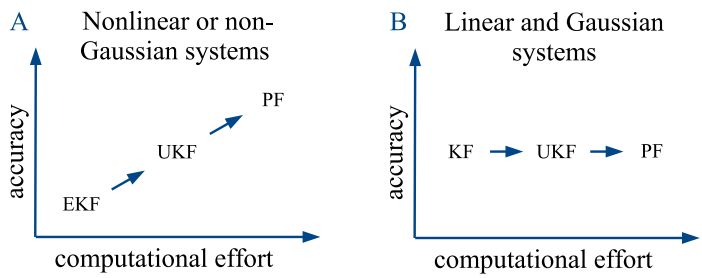


Figure 2.1. State estimation trade-offs (adapted from Simon, 2006, Fig. 15.7).

Parameter Estimation

In addition to estimating dynamic state variables, more static or slowly changing parameters may also be estimated online from the data using the estimation methods previously explained in this thesis (Simon, 2006, p. 422). The model of the dynamic system (i.e., functions f_k and h_k in Equation (2.4)) usually contains parameters which must be tuned such that the model represents the true world. If this is performed prior to the estimation, the process is called *calibration*. However, it is possible to estimate these parameters online, and then the process is called *parameter estimation*.

In order for the parameter vector $\mathbf{p}$ to be estimated, it must be observable, which means that there should be sufficient independent measurements to estimate all states and parameters. Then the state vector may be augmented with the parameter to obtain an augmented state vector $\mathbf{x}'$ (Simon, 2006, p. 423):

$$\mathbf{x}'_k = \begin{bmatrix} \mathbf{x}_k \\ \mathbf{p}_k \end{bmatrix} \tag{2.21}$$

If the parameter vector $\mathbf{p}_k$ is assumed to be constant over time indices k, then a small amount of artificial process noise $\boldsymbol{\nu}_{p,k}$ is required for the augmented system model to enable the filter to slowly tune the parameters. The larger the covariance of the noise is set in the filter, the faster the parameter is allowed to change in the filter. The dependency to the parameter vector and its noise is then added to the original system model in Equation (2.4) as follows:

$$\begin{aligned} \mathbf{x}'_k &= f_{k-1}\left(\mathbf{x}_{k-1}, \mathbf{p}_{k-1}, \mathbf{u}_{k-1}, \boldsymbol{\nu}_{k-1}, \boldsymbol{\nu}_{p,k-1}\right) \\ \mathbf{x}'_k &= f_{k-1}\left(\mathbf{x}'_{k-1}, \mathbf{u}_{k-1}, \boldsymbol{\nu}'_{k-1}\right), \end{aligned} \tag{2.22}$$

where $\mathbf{x}'_k$ is the augmented state, which includes the parameter vector $\mathbf{p}_k$, and $\boldsymbol{\nu}'_k$ is the augmented process noise, which includes the artificial noise parameter $\boldsymbol{\nu}_{p,k}$. If measurement model h_k is also dependent on the parameters, the measurement part of the state space model in Equation (2.4) must also be augmented:

$$\mathbf{y}_k = h_k\left(\mathbf{x}_k, \mathbf{p}_k, \boldsymbol{\upsilon}_k\right) = h_k\left(\mathbf{x}'_k, \boldsymbol{\upsilon}_k\right). \tag{2.23}$$

After augmentation, the parameter is modeled as a static state variable, and it can be estimated along with other state variables with any of the state estimation methods previously explained. For example, parameters are estimated in this manner in Publication II, where gyroscope biases are estimated along the attitude as an augmented state. Naturally, other-than-constant parameters may also be estimated if the dynamic model of the parameter is set accordingly in f_k in Equation (2.22).

Sensor Fusion

Usually, a single sensor is unable to gather sufficient relevant information from the robot and its environment. For example, the attitude of the robot is difficult to measure with any single sensor. Accelerometers or tilt sensors, which are commonly used for the task, do not measure linear accelerations alone but rather external specific force (i.e., proper acceleration, which is the acceleration relative to free-fall). Specific force is a combination of linear acceleration and the Earth's gravitational field (Hol, 2011, p. 25). Therefore, accelerometers and tilt sensors may be used in a static case to measure the direction of the Earth's gravitational field, but, when in motion, all other accelerations caused by the robot's own movement are summed to the sensor readings, causing an erroneous attitude estimate.

The obvious solution to the previous problem is to add more sensors. For example, gyroscopes are usually added to estimate changes in attitude while the accelerometer is measuring. This allows the estimator to average over noisy acceleration signals when fast rotations can be measured by the gyroscopes and compensated for in the sensor fusion algorithm. However, this raises new challenges for combining the data from these multiple sensors. These challenges relate to the time synchronization of the sensors, nonidealities in sensors, such as unknown or drifting calibration parameters, and balancing with real-time requirements, computational effort, and assumptions built into different algorithms. These challenges are studied under the name of *sensor fusion*. Sensor fusion is defined as the combination of sensory data, or data derived from sensory data, such that the resulting information is in some sense better than would be possible when these sources are used individually (Elmenreich, 2002, p. 8).

There exist multiple methods and architectures to fuse sensor measurements (see, e.g., Elmenreich, 2002, Ch. 3). However, the main principle is the same for all of them. In sensor fusion, measurements from multiple sensors are combined to produce a more accurate or enhanced estimate of reality than that offered by a single sensor. The main principle is visualized as a block diagram in Figure 2.2, where the environment and the robot's own state are measured with N sensors ($s_1 \ldots s_N$), and the measurements are combined in the sensor fusion block to recursively build a representation of the environment (e.g., a map) and the robot's own state in it. Together, this process is called *machine perception* (or alternatively robot perception), which usually aims to achieve some task (in Figure 2.2, the robot control block demonstrates a general task).

The same signal processing methods which were introduced in the previous few sections for state and parameter estimation can also be used for sensor fusion. Furthermore, most current sensor fusion methods are probabilistic, for example, Kalman filtering and Bayesian reasoning (Elmenreich, 2002).

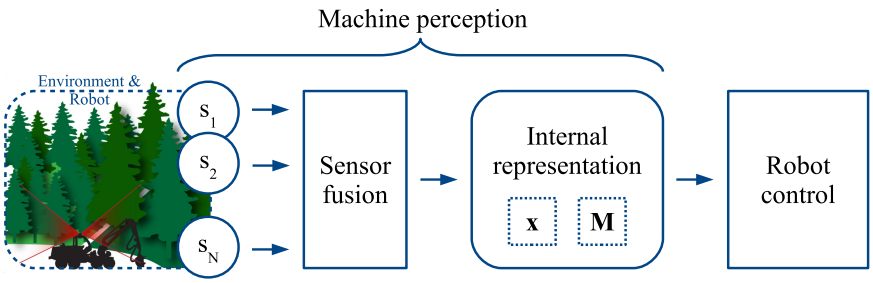


Figure 2.2. Block diagram of sensor fusion of sensors $s_i, i \in [1,N]$ to state **x** and map **M**

For sensor fusion to succeed, sensor data must be correctly associated and time synchronized, which means that measurements should be fused in the right order and the time differences between the data sources must be known and taken into account in the algorithm (Kaempchen & Dietmayer, 2003; Ding, Wang, Li, Mumford, & Rizos, 2008). This is because even a small time offset between the signals may result in a significant error (Hol, 2011, p. 78). Unfortunately, synchronization of different sensors is difficult, since most sensors exhibit nondeterministic measurement latencies, i.e., delays before the measurement arrives at the acquisition system (Huck, Westenberger, Fritzsche, Schwarz, & Dietmayer, 2011). Several factors contribute to this latency, including measurement acquisition time, pre-processing time, communication transfer, buffering, and computer scheduling, all of which occur before the measurement is received and time stamped in the software. These factors are usually subject to fluctuations and jitter, and they often cannot be directly measured (Huck et al., 2011).

In addition to timing, sensor models containing knowledge about the pose of the sensor and sensor inaccuracies or noise are important for fusing the signals. For example, in Publication VI, data from a camera and a lidar are fused together. If the mutual rotation and translation between the lidar and the camera are defined incorrectly, the area segmented based on the lidar sensor might not correspond to the area selected from the camera. This would lead to completely incorrect estimates, as the sensor fusion assumptions are not fulfilled. The parameters that relate to the pose of the sensor are usually called extrinsic (geometric) calibration parameters (e.g., Hanning, 2011, p. 14).

Furthermore, many sensors, such as lidars or cameras, feature complex measurement models which define how the sensor signals relate to the physical world. These models are often defined as a function of some parameters. In the computer vision literature, these parameters are usually referred to as intrinsic calibration parameters (e.g., Hanning, 2011, p. 14). Machine vision cameras usually also involve the use of parameters that model how the image is distorted. In addition, the intensity-related parameters of the imaging sensor may also be calibrated, which is known as

photometric calibration (e.g., Szeliski, 2010, Section 10.1).

Sensor calibration is required to identify the values for the defined calibration parameters. This is usually accomplished through an error minimization process which formulates the calibration error, measures some known object or shape in the environment, and tunes the parameters to minimize the error function (e.g., Hanning, 2011, Ch. 3). For example, in Publication I, a minimization procedure is formulated to tune lidar extrinsics, i.e., the position and orientation parameters of the scanner, and, in publications IV, V, and VI, a Matlab toolbox (Bouguet, 2004) is used for calibrating camera parameters. In addition, non-optical sensors usually contain calibration parameters. For example, the IMUs used in publications II–IV are calibrated using the methodology proposed in Publication II.

Sensor fusion may be categorized as low level, intermediate level, or high level depending on the amount of computation performed for the sensor readings before fusion (Elmenreich, 2002, p. 14). In high-level fusion, the sensors are usually loosely coupled. This form of fusion refers to a solution where the measurements from several individual sensors are pre-processed before they are used to compute the final result (Hol, 2011, p. 18). By contrast, lower-level fusion can be more tightly coupled, where the aim is to utilize all available information to improve the sensing. In a tightly coupled approach, all measurements are directly used to compute the final result (Hol, 2011, p. 18).

For example, Publication I uses deep integration of accelerometer and gyroscope measurements to enable gyroscope bias estimation alongside the main task of attitude estimation. In comparison, Publication IV proposes loosely coupled sensor fusion for integrating the attitude estimate from a head-mounted IMU and a camera to build a more reactive and robust head pose estimate than would be achieved with a head-mounted camera alone. In this high-level fusion, it is not possible to use the camera information to tune the IMU parameters or vice versa. However, higher-level fusion is simpler to implement.

In addition to fusion level, sensor fusion can be categorized into complementary, competitive, or cooperative fusion (Elmenreich, 2002, pp. 15–17):

a) Complementary fusion creates a spatially or temporally extended view (e.g., the back-to-back usage of two similar tilted laser scanners in Publication VII to double the scanning speed and data acquisition rate).

b) Competitive fusion provides robustness to a system by combining redundant information (e.g., IMU and camera fusion in Publication IV for estimating the orientation of the driver's head).

c) Cooperative fusion provides an emerging view of the environment by combining non-redundant information (e.g., the emergence of a

3D worldview by using a rotation sensor and a vertically mounted laser scanner in Publication I, and gyroscope bias estimation by combining the information from triaxial accelerometer and gyroscope measurements in Publication II).

In conclusion, sensor fusion can be and usually is accomplished with the same probabilistic tools that are used for state and parameter estimation. The potential advantages of sensor fusion are increased confidence, reduced ambiguity and uncertainty, robustness against interference and faults, extended spatial and temporal coverage, and improved resolution (Elmenreich, 2002). In addition, cooperative sensor fusion is able to produce novel capabilities that are not possible through the use of any one sensor alone.

2.2 Attitude Determination and Positioning

Attitude is usually measured with inclinometers and a compass or estimated with the help of digital sensors, such as magnetometers and inertial measurement units (IMUs). An IMU is an instrument that is built from triaxial accelerometers and gyroscopes that sense accelerating forces (including gravity) and rotation rates in three orthogonal axes relative to an inertial reference frame (Petovello, 2003). Traditionally, these sensors have been highly expensive mechanical or optical devices (Tazartes, 2014), but low-cost solutions have recently emerged in the form of micro-electro-mechanical systems (MEMS), which are integrated chips containing miniature mechanical constructions packaged together on a silicon chip with measurement and sampling electronics (Aggarwal, 2010).

Magnetometers, which sense the direction and magnitude of the magnetic field, are added to the setup to allow heading estimation by assuming that the magnetic field they measure is mostly caused by the Earth's magnetic field. IMUs and magnetometers can be used together to construct an attitude and heading reference system (AHRS). Such systems are sometimes called magnetic angular rate and gravity (MARG) sensor arrays (Madgwick et al., 2011). In many cases, the Earth's magnetic field is a reliable indicator of the direction of magnetic north. However, forest machines are usually built from ferromagnetic steel, and they contain moving parts and changing electric current flows. They can also operate under power lines, which may induce magnetic disturbances. Magnetometer measurements are highly susceptible to magnetic disturbances, and, in particular, the magnetic field generated by ambient ferromagnetic materials can be more prominent than the geomagnetic field (Majumder & Deen, 2020). Thus, these sensors are unreliable and seldom used in forest machinery.

For static cases, a triaxial accelerometer can be used to measure the inclination of the sensor in the Earth's gravitational field and thus compute

the attitude of the sensor (except the heading) (e.g., Luczak, Oleksiuk, & Bodnicki, 2006). However, MEMS accelerometers are often noisy[14]. Therefore, several accelerometer measurements must be averaged together to estimate the attitude reliably. This leads to problems in dynamic cases, where the sensor may be rotating while measuring. Therefore, the rotation of the sensor between the measurements should be known in order to combine accelerometer measurements taken at different times.

Gyroscopes, which measure the rotation rates of a rigid body, have been added to the setup to overcome the previous problem and to enable dynamic measurements. However, as found in the previous sensor-fusion section, adding more sensors might not simplify the solution. Low-cost MEMS sensors are noisy, their measurements include errors, and the calibration parameters might change, which makes attitude estimation a complex task. These parameters consist of an unknown zero level, i.e., a *bias*, and an unknown scale factor, i.e., a *gain* (Lou, Xu, Cao, Chen, & Xu, 2011). In addition, the gain and the bias tend to drift over time and are typically temperature dependent (Hol, 2011, p. 25). For these reasons, sensor calibration is one of the most challenging issues in inertial navigation (Sahawneh & Jarrah, 2008).

Furthermore, if the sensor is moving or accelerating in addition to rotating, the accelerometers measure all linear accelerations caused by the motion and they are summed to the gravity measurement. As mentioned earlier, an accelerometer measures the external specific force (combined accelerations and gravity) (Hol, 2011, p. 25). As gravity and linear accelerations are indistinguishable in the measurements, the attitude estimation process becomes significantly more difficult in unrestricted dynamic cases.

Attitude estimation is a classic problem in the field of sensor fusion, and there are plenty of different algorithms available (see, e.g., the survey by Crassidis, Markley, & Cheng, 2007). Many of the existing algorithms have traditionally relied on data obtained from military grade IMUs, which usually involve export restrictions and high cost, which limits commercial applications (Baldwin, Mahony, Trumpf, Hamel, & Cheviron, 2007). Newer methods have mostly been developed for lower-cost MEMS devices. However, these cheaper commercial-grade IMUs commonly experience high levels of non-Gaussian noise in their gyroscope and accelerometer measurements, which often leads to the instability of classical Kalman and extended Kalman filter algorithms (Baldwin et al., 2007; Crassidis et al., 2007).

IMU and AHRS sensor fusion algorithms are usually Kalman filter solutions (e.g., Bistrov, 2012; Ercan et al., 2011; Lou et al., 2011; Phuong, Kang, Suh, & Ro, 2009; Jurman, Jankovec, Kamnik, & Topič, 2007) or extended Kalman filters (e.g., Luinge & Veltink, 2004; Barshan & Durrant-

[14] E.g., the low-cost MEMS IMU sensor MPU-9250 manufactured by InvenSense produces root mean square (RMS) noise of about 0.08 m/s^2 (InvenSense, 2016)

Whyte, 1995; Foxlin, 1996; Edwan, Zhang, Zhou, & Loffeld, 2011; Gebre-Egziabher, Hayward, & Powell, 2004), but there exist some non-Kalman filter solutions as well. These are complementary filters (e.g., Euston, Coote, Mahony, Kim, & Hamel, 2008), a gradient descent minimization for an error function by Madgwick et al. (2011), algorithmic heuristics which balance between calibration and a combination of different attitude sources (e.g., Zhou, Li, & Shen, 2014), or filters derived in Lie algebra for special Euclidean (e.g., Baldwin et al., 2007) or orthogonal groups (e.g., Mahony et al., 2008).

Since low-cost MEMS contain errors that develop with respect to time (e.g., drifting gyroscope biases), attitude estimation algorithms should include an online bias estimator for gyroscope biases. Unfortunately, only a minority of previously published algorithms include such an estimator. In many of these, other measurements are required in addition to gyroscopes and accelerometers to estimate gyroscope biases. The most commonly used sensor is a triaxial magnetometer, which is utilized in (Lou et al., 2011; Foxlin, 1996; Edwan et al., 2011; Gebre-Egziabher et al., 2004; Madgwick et al., 2011). In addition, satellite navigation (Lou et al., 2011; Gebre-Egziabher et al., 2004) or tilt sensors (Barshan & Durrant-Whyte, 1995) are used for bias estimation. Other innovative solutions also exist, such as the combination of two IMUs in a controlled inclination offset (Ruizenaar, van der Hall, & Weiss, 2013) and the mathematical decoupling of magnetometer measurements to prevent them from affecting attitude or bias estimation (Hua, Ducard, Hamel, Mahony, & Rudin, 2013).

In addition to Publication II, only a small number of algorithms (Hamel & Mahony, 2006; Mahony et al., 2008; Wu, Sun, Zhang, & Chen, 2014) have been proposed to estimate gyroscope biases without any extra sensors in addition to the triaxial accelerometer and gyroscope. To the best of the author's knowledge, Publication II is the only such method that provides opensource code (Hyyti, 2015).

Position is usually measured either locally using dead reckoning or odometry methods to acquire a relative local position (Mohamed et al., 2019) or globally using a global satellite navigation system (GNSS), e.g., the U.S.-owned Global Positioning System (GPS), the Russian GLONASS, or the European Galileo system (Hofmann-Wellenhof, Lichtenegger, & Wasle, 2007). These satellite navigation systems provide geolocation and time information to a receiver anywhere on Earth as long as there is an unobstructed line of sight to four or more satellites of the same system (Campbell et al., 2018).

According to Mohamed et al. (2019), relative local position can be measured with five different odometry methods: wheel, inertial, radar, visual, and laser odometries (also called lidar odometry). Odometry traditionally refers to the use of measurements from actuators (wheels, treads, etc.) to estimate vehicle motion (Siciliano & Khatib, 2016, p. 737). Since there

are also other means to perform positioning from the measurements, this simplest and oldest[15] form of self-contained localization is referred to as *wheel odometry* (Mohamed et al., 2019). In wheel odometry, the rotation and steering angle of wheels is usually measured with encoders, and this motion is integrated into a dynamic model to determine the vehicle's current position relative to the starting point (Mohamed et al., 2019; Dudek & Jenkin, 2010, pp. 44–46).

In addition, inertial navigation systems (INS) are used for positioning (sometimes referred to as inertial odometry (e.g., Mohamed et al., 2019)). INS systems usually contain more precise inertial sensors than the IMUs used and referred to in this work. An INS integrates rotation rates to obtain orientation changes and doubly integrates gravity-reduced accelerations to obtain velocity and position increments (Jekeli, 2000; Petovello, 2003). Unfortunately, the double integration of accelerations is prone to errors, since accelerometer and gyroscope measurements contain noise and nonidealities, such as a temperature dependent non-zero offset (i.e., bias) in the measurements.

Remote sensors such as radars, lidars, and optical cameras are also used to estimate vehicle motion. All of the related odometry methods (radar, lidar, and visual odometry) share the same main concept (Mohamed et al., 2019). If the surrounding environment can be assumed to be static, all the changes in the observations are caused by the observer's own motion. In radar odometry, radio signals are emitted and received to measure the velocity and range of objects around the vehicle (Mohamed et al., 2019). Then, suitable features are extracted and tracked in the measurements over time. In radar odometry, the Doppler shift can also be utilized to estimate the velocity of the vehicle (e.g., Vivet, Checchin, & Chapuis, 2013; Kellner, Barjenbruch, Klappstein, Dickmann, & Dietmayer, 2014).

In lidar odometry, two adjacent scans (i.e., observed point clouds in two consecutive lidar sweeps) are compared to each other using a method which computes a transformation between two point clouds (Mohamed et al., 2019). Commonly, iterative closest point (ICP) variants have been used, which minimize distances between the nearest points between the point clouds. The ICP uses an initial guess for the point cloud's relative rigid-body transform, and it iteratively refines the transform by repeatedly generating pairs of corresponding points by minimizing the error metric (Rusinkiewicz & Levoy, 2001). In addition, point-to-plane (Low, 2004), point-to-line (Censi, 2008), and generalized (Segal, Haehnel, & Thrun, 2009) error metrics have been used. Instead of comparing random points between point clouds, other proposed methods have used computed features from the point clouds, such as normal distributions (Biber & Straßer, 2003; Magnusson, 2009; Stoyanov, 2012) or, more recently, line and corner

[15] The basic concepts that underly odometry have been studied for more than two millennia (Siciliano & Khatib, 2016, p. 737).

features (J. Zhang & Singh, 2014, 2017).

In visual odometry, the position and orientation of a platform is estimated by analyzing the variations induced by the motion of a camera (or two or more cameras together) on a sequence of images (Mohamed et al., 2019). Existing visual odometry algorithms differ according to whether they have been designed for a monocular camera (e.g., Forster, Pizzoli, & Scaramuzza, 2014) or a stereo camera pair (e.g., Howard, 2008). Furthermore, methods for depth cameras (e.g., Whelan, Johannsson, Kaess, Leonard, & McDonald, 2013) and omnidirectional (e.g., Scaramuzza & Siegwart, 2008) and fisheye (e.g., P. Hansen, Alismail, Rander, & Browning, 2013) cameras use different methods that are specially designed for the different camera type. In addition, methods also differ according to the key information upon which odometry is performed. This can be raw pixel values that are matched together, for instance, using a template matching algorithm (e.g., Nourani-Vatani & Borges, 2011). More sophisticated features (e.g., SIFT by Lowe, 2004) can also be sought from the images to track ego-motion over time (Chien, Chuang, Chen, & Klette, 2016).

Positioning sensors and techniques may also be integrated to gain a more reliable and accurate result. High-accuracy tight-integrated GNSS/INS systems can provide accurate location and orientation information even if satellites are not always visible (see Petovello, 2003, ch. 9). In the forest, the accuracy of satellite navigation under forest foliage is a limiting factor for positioning, as there rarely is an unobstructed line of sight to the satellites. However, Kaartinen et al. (2015) have shown that a high-end, tightly-integrated inertial navigation and GNSS receiver system can achieve 0.7-meter accuracy under forest canopies by using post-processing methods. These methods utilize a reference base station at a known location and generally result in a more accurate, comprehensive solution than is possible in real-time (Novatel, 2015). These reference stations are used to model satellite-specific biases, such as clock bias and orbital errors, and signal propagation-medium-related biases caused by, for example, ionospheric and tropospheric refraction (Hofmann-Wellenhof et al., 2007, p. 109). Furthermore, the reference station commonly calculates correction data, which can also be transmitted to the remote receiver in real time to improve positioning accuracy (Hofmann-Wellenhof et al., 2007, p. 169).

Similar to GNSS sensors, radars, lidars, and cameras can also be integrated with inertial sensors to improve accuracy and robustness (Mohamed et al., 2019). These sensor fusion methods are called radar-inertial odometry (RIO) (e.g., Doer & Trommer, 2020), lidar-inertial odometry (LIO) (e.g., Ye, Chen, & Liu, 2019; T. Shan et al., 2020), and visual-inertial odometry (VIO) (e.g., M. Li & Mourikis, 2013). In addition, many modern positioning methods utilize or build a map to improve positioning accuracy. These methods are studied in the field of simultaneous localization and mapping (SLAM) (e.g., Spero & Jarvis, 2007; Lu et al., 2009; Khairuddin et al.,

2015; Karam, Lehtola, & Vosselman, 2020; Taheri & Xia, 2021). As an example, positioning accuracy in a forest can be improved to a few centimeters by using SLAM-based methods in combination with a high-end, tightly-integrated inertial navigation and GNSS receiver system, as shown by Kukko et al. (2017).

In SLAM, the robot builds a map of the environment consisting of landmarks and possibly other features, such as obstacles or topography, and the same map is also used for defining the robot's location in the world (Spero & Jarvis, 2007). In the beginning, if the robot lacks a priori knowledge, it is at an unknown location in an a priori unknown environment. It then uses its sensors to observe the nearby environment to detect suitable landmarks. It also measures its own pose with respect to the detected landmark locations. As the robot moves through the environment, its changing viewpoint enables incremental map building from these landmarks. The generated map is continuously utilized to track the robot's current pose relative to its original pose where the mapping began (Spero & Jarvis, 2007).

If the robot already possesses a map but the initial location on the map is unknown, the problem is called a global positioning problem (also known as a wake-up robot problem) (Fox, Burgard, & Thrun, 1999). Such a problem is more difficult than the position-tracking problem normally solved in SLAM (Thrun et al., 2005, p. 159). When solving the global positioning problem, the robot must be able to deal with multiple possible hypotheses about its pose, since a similar set of landmarks might be situated in multiple locations and orientations around the map.

Nonlinear filtering methods such as particle filters (introduced in the latter part of Section 2.1) are better at dealing with the multiple hypotheses required in the global positioning problem. Particle-filter-based localization methods can also be injected with random initial guesses to compensate for an erroneous global position estimate (see, e.g., Thrun et al., 2005, Sec. 8.3.3, pp. 204–209). Instead of injecting the particle filter with random particles, measurements from another system can also be added into a particle filter. Then, position measurements from another positioning system are induced as particles in the particle filter instead of randomly assigned particles. In such a case, the estimate in the particle filter is slightly biased towards the measurements acquired from the other positioning system. This kind of augmentation is used in Publication I to recover from errors.

2.3 Crane and Tool Pose Measurement

In robotics, the pose of the end effector is traditionally measured from inside the robot using an angular position sensor (e.g., a potentiometer, optical encoder, or tachometer) for each joint angle (Mihelj et al., 2019, p. 86). For hydraulic manipulators, such as forest machines, joint angles

can also be measured using a linear position sensor attached to each hydraulic cylinder, and telescopic extensions (i.e., prismatic joints) common in forest machines can be measured with a length-measuring device built within the joint (Kalmari et al., 2013; Lindroos et al., 2015).

When the cylinder lengths are known, joint angles can be calculated using kinematic equations, and the end effector position can then be calculated using a forward kinematic chain of rigid links (Waldron & Schmiedeler, 2008). This assumes that the crane parts are rigid structures that do not bend under stress. Instead of using the stiff structures more familiar in industrial robotics, hydraulic forestry cranes are usually built to be flexible to optimize material usage and keep their weight low (Pedersen, Andersen, & Nielsen, 2015). This results in significant bending, which should be taken into account when the crane posture and end effector position are estimated.

Methods to estimate the bending of the manipulator (e.g., De Luca & Panzieri, 1994) could be used to cope with the problem. However, bending-estimation methods must use the right model parameters and predefined weights to avoid end-effector displacement errors. In addition, these weights are bound to change as the crane is used to lift unknown loads, for instance, when handling various size logs, felling trees, or exerting unknown forces when uprooting vegetation. Therefore, there is a need for alternative methods to maintain a sufficient level of accuracy for estimating the end-effector position or the crane posture. Such alternative methods include observing the pose by using optical sensors, such as cameras and laser scanners, or, alternatively, by using inertial measurement units (IMUs).

Camera-based solutions have previously been demonstrated mostly for excavators (e.g., Mulligan, Mackworth, & Lawrence, 1989; Mielikäinen, Koskinen, Handroos, Toivanen, & Kälviäinen, 2001; Feng, Dong, Lundeen, Xiao, & Kamat, 2015), for large tower cranes (e.g., Yang, Vela, Teizer, & Shi, 2012), for large rope-operated shovels (e.g., Corke, Roberts, & Winstanley, 1998; Lin, Lawrence, & Hall, 2010), and for underground mining machinery (e.g., Corke et al., 1998). In addition, lidar-based solutions have focused on estimating the pose of large mining machinery, such as large rope-operated shovels (e.g., Dunbabin & Corke, 2006; Kashani, Owen, Himmelman, Lawrence, & Hall, 2010; Phillips, Green, & McAree, 2016). Only Kashani, Owen, Lawrence, and Hall (2007) have tested a laser scanner with a normal-sized excavator. To the author's knowledge, the method proposed in Publication I is the only forest-machine crane-posture measurement solution using machine vision or laser scanning.

Inertial measurement-based solutions usually employ one or more IMUs to estimate the orientation of the sensor with respect to the Earth frame (i.e., attitude). Then the attitude of each IMU is used to measure the orientation of the crane part on which the sensor is mounted. In research

by Vihonen et al., (2013a, 2013b, 2014; 2016) the pose of a forestry crane is estimated with a group of low-cost MEMS IMUs. Their latest work includes an integrated kinematic model for the crane using three IMUs, one attached to each boom segment (Vihonen et al., 2016). The setup can reach up to 1 degree accuracy for lift and transfer (referred to as tilt in their work) angles. In addition to the many studies by Vihonen et al. (2013a, 2013b, 2014, 2016), the DCM IMU algorithm (Hyyti, 2015) (proposed in Publication II) was also successfully employed with a kinematic crane model (see Publication I) to estimate the posture of a forestry crane in the master's thesis of Toiviainen (2017).

Inertial measurements can be accurate for attitude estimation (as shown in Publication II). However, the length of a telescopic extension is significantly more difficult to measure or estimate with IMUs alone. In their study, Vihonen et al. (2014) could achieve centimeter-level accuracy when the telescopic link was contracted, but, while the crane was extended at its maximum reach, the average error increased to 0.26 meters. This is significantly more than the average error present in Publication I. For this reason, extra range measuring instruments are preferred for measuring the length of the prismatic joint. For example, P. Cheng, Oelmann, and Linnarsson (2011) extend IMU instrumentation with a sonar sensor to measure the length of the extension link.

In addition to estimating the pose of the crane, IMUs have also proven useful for estimating the pose of a freely swaying tool attached to the boom tip, as shown in Publication III. The three-axis rotation of the tool relative to the crane or boom tip pose is difficult to measure using traditional sensors (e.g., encoders and position sensors), since the tool is suspended with a rotator link and a rotator (see Figure 2.3). This means that the tool freely hangs from the tip of the boom with a rotator link that allows free swaying in two opposing directions. The tool can be rotated with a hydraulic motor called a rotator, which is mounted after the link.

Installing sensors on each of these joints is challenging due to the mechanical structure of the rotator link. In addition, the tool is often in contact with obstacles in the forest, and thus the sensors should be well protected or built inside the mechanism. To the best of the author's knowledge, no studies have proposed methods for the measurement of the freely hanging rotator link and the rotator motor except Publication III. The method proposed in Publication III is able to estimate the 3D angular position and angular velocities of the freely hanging tool with an accuracy of a few degrees.

In addition to optical and inertial measurements, ultra wide-band radio frequency identification (UWB RFID) tags (C. Zhang, Hammad, & Rodriguez, 2011) and global navigation satellite system (GNSS) receivers (Kim & Langley, 2003) have also been used to measure crane pose. C. Zhang et al. (2011) studied the UWB positioning of a construction crane for safety

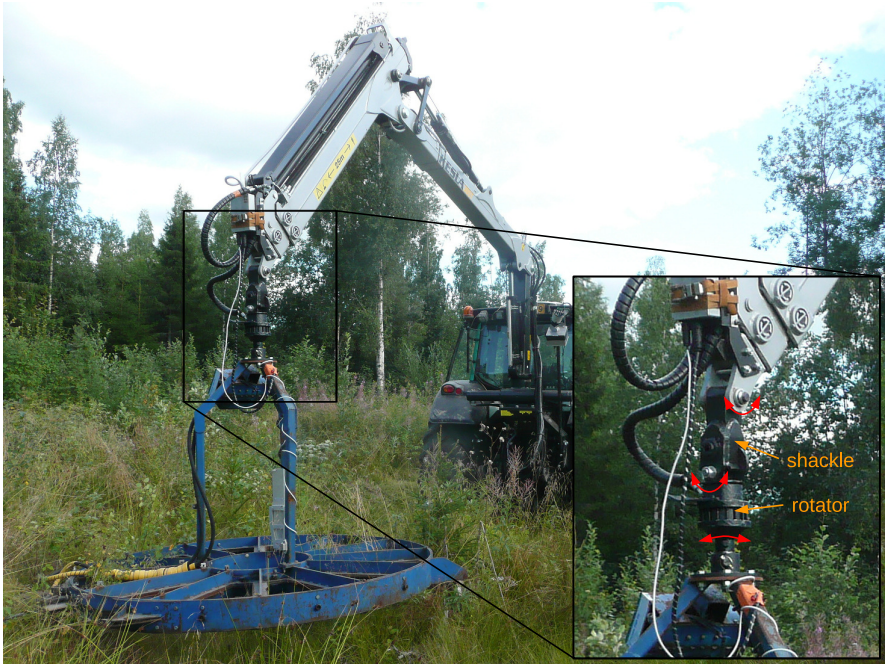


Figure 2.3. A blue cleaning tool is mounted on the forestry crane using a rotator link and a rotator with three rotation axes (red arrows), from which the top two on the rotator link swing freely, and the lowest one is rotated by a hydraulic motor called a rotator.

purposes on an open construction site, obtaining a position accuracy of approximately 25 cm with active RFID tags. In their work, transceivers were installed in the environment to guarantee good visibility between them and the tags. UWB techniques have since been shown to yield sub-centimeter accuracies in an indoor-positioning competition (see Table 2 in Lymberopoulos & Liu, 2017). Therefore, UWB technologies could also possess potential for crane posture estimation in the future.

2.4 Measuring the Surrounding Forest

In autonomous vehicles, exteroceptive senses are usually built using sensors such as machine vision cameras, radars, sonars (including ultrasonic sensors), and lidars (Campbell et al., 2018). To be usable in a forest environment, these sensors must tolerate outdoor conditions, for instance natural light, shadows, fog, dust, rain, wind, snow, and normal boreal seasonal temperatures ranging from hot summers to freezing winters (Vestlund & Hellström, 2006).

Camera

Machine vision is a low-cost, fast, and powerful sensing method to collect a large amount of high-resolution data (Hague, Marchant, & Tillett, 2000). The main advantage of machine vision cameras over other remote sensing equipment is their ability to see colors and textures (Campbell et al., 2018). For example, machine vision cameras have been used inside the forest for log length measurement in a harvester head (Kalmari et al., 2011), annual ring width measurement (Marjanen, Ojala, & Ihalainen, 2008), forest structure estimation (Kulovesi, 2009), tree stem detection (Hellström & Ostovar, 2014), stem damage detection (Palander et al., 2019), tree species identification (Carpentier, Giguere, & Gaudreault, 2018), forest machine orientation estimation (Matej, 2014) and navigation (Schulze, 2012), to name but a few applications.

A color camera measures the intensity of light illuminated on individual pixels of the sensor array. The sensor array features a different spectral response function in its different color sensors, from which the most common is trichromatic red, green, and blue (RGB) (Szeliski, 2010, pp. 75–76). Color cameras are often produced by using a color filter array (e.g., a Bayer filter by Bayer, 1976) attached on top of a single photosensitive image sensor such that each pixel has its own dedicated filter (Sonka, Hlavac, & Boyle, 2014, p. 44). This means that each pixel measures only one color, and the neighboring pixel values must be used to interpolate values for the other colors using a demosaicing algorithm (X. Li, Gunturk, & Zhang, 2008). In addition, different configurations exist which employ a prism to split the incoming light into multiple different image sensors that are made sensitive to different wavelengths; alternatively, a filter can be changed between consecutive images on the same image sensor (Sonka et al., 2014, p. 44). In addition, tunable filters such as Fabry-Perot interferometers have been used for hyper-spectral imaging (e.g., Rissanen et al., 2017).

In RGB, an image is represented as a vector of the three primary colors (red, green, and blue) for each pixel. There also exist various different color models, i.e., digital representations of colors (Ibraheem, Hasan, Khan, & Mishra, 2012). Some are highly device-dependent, like RGB, which contains the intensities measured at these three spectral channels. RGB is a poor choice for color image processing, since it is highly correlated among channels, it mixes chrominance (i.e., the color information) and luminance (i.e., intensity information), and it is nonlinear to perceptual visual observation (Saravanan, Yamuna, & Nandhini, 2016). Instead, other color models that are less correlated, such as hue-saturation-lightness value / intensity / luminance (HSV/HSI/HSL) color spaces could be used instead. These are nonlinear conversions of the RGB color space that aim to separate the luminance part from the chrominance data, which makes it beneficial for image processing (Saravanan et al., 2016). The same

color transformation has also proven usable for visualizing point clouds (e.g., Publication I). However, one of the disadvantages of this nonlinear transformation is that the hue contains a non-removable singularity near the axis of the color cylinder (H.-D. Cheng, Jiang, Sun, & Wang, 2001).

There also exist linear color model transformations from RGB, which can be extremely effective alternatives to nonlinear conversions that avoid singularities but still reduce the correlation between color channels. The best color features for natural image segmentation have been found by using a 3D projection on an RGB color space using a linear transformation to excessive green (EG), redness-blueness (RB), and intensity (I) channels (Ohta, Kanade, & Sakai, 1980; Steward & Tian, 1999):

$$\begin{bmatrix} EG \\ RB \\ I \end{bmatrix} = \begin{bmatrix} -\frac{1}{2} & 1 & -\frac{1}{2} \\ 1 & 0 & -1 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{2.24}$$

Ohta et al. (1980) compared multiple different color transformations for segmentation purposes and found that the fixed linear transformation defined in Equation (2.24) was the best approximation at each important step in segmenting natural images. Furthermore, as it separates image intensity (I channel), it is easy to reduce the effect of shadows by normalizing intensity in images by dividing each EG and RB channel pixel-wise by the intensity channel, as in Publication V.

In addition to color, machine vision is usually used to detect objects based on their texture, which refers to properties that represent the surface or structure of an object (Sonka et al., 2014, p. 747). Although texture is widely used and mostly intuitively understood, it nonetheless lacks a precise definition (Sonka et al., 2014, p. 777). In images, texture generally refers to the repetition of basic texture elements, i.e., texture primites called texels (Patel & Tandel, 2016; Sonka et al., 2014, p. 777). These texels contain several pixels, whose placement can be periodic or random. In natural environments, such as forests, textures are generally random, whereas artificial textures are often deterministic or periodic (Patel & Tandel, 2016). As an example of an artificial texture, an ArUco marker is a specially crafted black and white marker that includes a code for the computer vision to detect it, measure its orientation, and distinguish individual codes (Garrido-Jurado, Muñoz-Salinas, Madrid-Cuevas, & Marín-Jiménez, 2014). In forest-related machine vision research, texture is mostly used for object segmentation and detection (e.g., Ali, Georgsson, & Hellstrom, 2008; Hellström & Ostovar, 2014).

There are abundant texture feature extraction methods in the literature. Commonly, spatial frequencies are measured in the image (Sonka et al., 2014, p. 750). In addition, edges can be detected or gradients computed prior to texture feature extraction (Sonka et al., 2014, p. 754). A review by Humeau-Heurtier (2019) categorizes texture feature extraction methods

into seven classes: statistical, structural, transform-based, model-based, graph-based, learning-based, and entropy-based approaches. From these, the first four are already well established; however, structural approaches are unsuitable for the highly random textures present in natural imagery (Humeau-Heurtier, 2019).

In statistical approaches, data on the spatial distribution of gray levels are used as texture descriptors. Many different methods have been developed for this task. Such methods include a gray-level co-occurrence matrix (GLCM), which computes how often pairs of pixels with a specific value and offset occur in the image, a local binary pattern (LBP), which codes each pixel as an 8-bit sequence based on its neighboring pixels' intensities, an autocorrelation function (ACF), which defines the similarity of an image patch to a shifted version of itself, and histogram based methods (Humeau-Heurtier, 2019; Ramola, Shakya, & Van Pham, 2020). According to Ramola et al. (2020), GLCM is the best approach for analyzing surface texture and landcover classification for satellite data processing, while LBP is widely used to analyze individual facial features, and autocorrelation is employed to identify the regularity of the textured surface. Humeau-Heurtier (2019) note that a histogram of gradient magnitudes often outperforms LBP-based methods because a histogram can be built to be rotation invariant, and the method has low computational complexity.

The main limitation of using texture as a measure of surface structure is related to how a camera measures the target. The structure of an object is not directly observed by the camera. Instead, in the camera image, texture is expressed by the spatial arrangement of color or intensity of adjacent pixels. For example, the differences between neighboring pixels provide us with information about local differences in the image, which are related to the actual differences on the observed surface. It is important to note that camera resolution and distance from the object change how the camera measures texture. The same surface may appear completely different at close range than when observed from a distance. This means that texture description is scale dependent (Sonka et al., 2014, p. 747).

To reduce the sensitivity to scale, a texture may be described in multiple resolutions and an appropriate scale chosen to achieve maximal discrimination in the classification task (Sonka et al., 2014, p. 764). Usually, some kind of transformation is also applied to the image before the texture detection or classification process. Consequently, Humeau-Heurtier (2019) terms these methods transform-based approaches, which represent an image in a space (e.g., the frequency or the scale space), whose coordinate system is better suited to analyzing the characteristics of texture. One commonly used transformation is the Fourier transformation[16], which brings spatial frequency information into focus. Alternatively, a bank

[16]This transformation is named after Joseph Fourier (1822), who demonstrated that some functions could be written as an infinite sum of harmonics.

of filters, such as various wavelet transformations, have also been used. Wavelet-based texture-description methods are often utilized, since they are typically more efficient than other statistical methods (Sonka et al., 2014, p. 778). Wavelets are short wave-like oscillations with an amplitude that begins at zero, increases, and then decreases back to zero, and with a zero average (Mallat, 1999, p. 72).

Similar to the Fourier transformation, which decomposes the signal into a family of complex sinusoids, the *wavelet transform* decomposes a signal into a family of wavelets. Contrary to the symmetric, smooth, and regular sinusoids in a Fourier transformation, wavelet-based methods define a sparse representation of piecewise regular signals, which may include transients and singularities (Mallat, 1999, p. 2). In images, large wavelet coefficients are usually located in the neighborhood of edges and irregular textures.

The discrete wavelet transform (DWT) is analogous to altering the input signal with a bank of bandpass filters, whose impulse responses are all approximately given by scaled versions of a "mother" wavelet (Kingsbury & Magarey, 1998). The scaling factor between adjacent filters is usually 2:1, which leads to octave bandwidths between outputs. Discrete wavelet frames can be designed to be invariant to translation (i.e., invariance to any circular shift in the input image), and as the input images are also discrete, they can be efficiently implemented using precomputed filter banks (Sonka et al., 2014, p. 778).

The selection of suitable wavelets is difficult, and usually some ready-made dictionary of wavelets is used. The simplest and earliest solution was the Haar wavelet by Alfred Haar (1910). He found that an arbitrary signal may be considered a linear combination of appropriately selected rectangular pulses. These pulses may be positive or negative and of suitable widths and heights (Sundararajan, 2015, p. 97). This odd rectangular pulse pair is the simplest possible wavelet (Stanković & Falkowski, 2003).

Gabor wavelets are a common, more complex set of wavelets invented by Dennis Gabor using complex functions constructed to serve as the basis for Fourier transforms (Gabor, 1946). Gabor found that Gaussian-modulated complex exponentials provide the best trade-off between time and frequency resolution (Lee, 1996). Gabor wavelets are also an interesting set for natural image processing tasks because the primary visual cortex in primates has been found to decode visual information similarly (Daugman, 1980; J. P. Jones & Palmer, 1987). According to Kamarainen (2012), such wavelets are also among the top performers in many technical applications, such as face detection and recognition, iris recognition, and fingerprint matching.

The transformation approach can also offer other advantages if the transformed space contains assumptions that benefit the detection of the target signal. For example, if the target has a known symmetry, this can

be taken into account by transforming into a suitable space. Trees, for instance, are usually rotation symmetric when viewed from above. This might considerably simplify detection, as there is no need to take account of image rotation if a rotation-invariant method can be used.

One possible solution for achieving rotation invariance is to perform a Radon transform prior to a translation-invariant wavelet transformation (Jafari-Khouzani & Soltanian-Zadeh, 2005). The Radon transform[17] takes a line integral over a line projected on 2D data. It is analogous to the Hough transform[18], which was originally used to detect lines on an image (van Ginkel, Hendriks, & van Vliet, 2004). These transformations project 2D image data into a space where translation on the x axis is related to an angle of the projected line and translation on the y axis relates to the distance of the projected line from the origin. Therefore, translation invariance introduced by the wavelet transformation can be changed to rotation invariance using a Radon transformation as proposed by Jafari-Khouzani and Soltanian-Zadeh (2005) and also as attempted in Publication V.

Alternatively, the output of a 2D wavelet transformation (i.e., separate output blocks in vertical and horizontal directions) may be combined as a single output such that the output is forced to be invariant for rotations. This kind of solution is proposed by Porter and Canagarajah (1997a) and is tested, among other features, in Publication VI.

These texture and color features are usually used as input data for classifying either the whole image or parts of the image. The latter is usually referred to as segmentation, where the image is split into multiple separate segments. These classifiers are usually defined as either parametric or nonparametric, although semiparametric solutions also exist (Lampinen, Laaksonen, & Oja, 1998, p. 22). A classical parametric approach is to model the class-conditional densities as multivariate Gaussians (Lampinen et al., 1998, p. 23).

One classical nonparametric method is a k-nearest neighborhood (k-NN) classifier[19], in which each class is represented by a set of prototype vectors. A new sample is assigned to the class most commonly represented in the collection of its k nearest neighbors (Lampinen et al., 1998, p. 31). In classical pattern recognition, this nonparametric k-NN classification method has been popular for many decades. According to Lampinen et al. (1998, p. 31), it can even be regarded as a sort of a baseline classifier against which other classifiers should be compared. The k-NN classifier is used to classify spruce among other vegetation using texture and color features in Publication V.

[17] The transform was introduced in 1917 by Johann Radon (1986).

[18] Invented by Paul V. C. Hough (1959) for detecting line segments in bubble chamber pictures.

[19] Fix and Hodges introduced a non-parametric method for pattern classification in 1951 that has since become known as the k-nearest neighbor rule (later published in Fix & Hodges, 1989).

A naive Bayes classifier (Rish, 2001) computes the probability density function of a class label L given the feature vector $\mathbf{f} = [F_1, F_2, ..., F_n]$. The naivety comes from the assumption that features are independent of each other, i.e., $p(F_i) = p(F), \forall i$. This simplifies the Bayesian posterior model to

$$p(L \mid \mathbf{f}) = \frac{p(L)}{p(F)} \prod_{i=1}^{n} p(F_i \mid L), \tag{2.25}$$

where n is the number of features, $p(L)$ describes the prior knowledge about the prevalence of each class L, and $p(F)$ is a normalizing term. A two-class naive Bayes classifier is used in Publication VI to classify spruce and birch tree segments on images.

Machine vision has drawbacks, especially outdoors, where it must operate in natural lighting conditions such as direct sunlight, shadows, and nighttime low-light intensity (Vestlund & Hellström, 2006). Furthermore, all sensors that use light on the visible spectrum may be impeded by adverse weather conditions, such as snow, fog, or heavy rain (e.g., Yoneda, Suganuma, Yanase, & Aldibaja, 2019).

Another disadvantage of machine vision is that the camera does not natively measure the range to the target, and therefore images are only 2D projections of the scene. However, there exist methods to circumvent this limitation. In a calibrated camera, coordinates of one image point define a ray in space uniquely. If two calibrated cameras observe the same point at the same time, its 3D coordinates can be computed as the intersection of two such rays (Sonka et al., 2014, p. 609). This is the basic principle of stereo vision. If only one camera can be used, then camera motion can be utilized to build a virtual stereo over a time interval. If the two poses of the same camera are known, the distance to the same static object can be triangulated (Sonka et al., 2014, p. 603). If the poses are not known, the problem becomes considerably more difficult, since the camera poses must also be estimated. This kind of multiple view geometry problem is usually solved using a non-linear least squares minimization method, known from photogrammetry as bundle adjustment (Sonka et al., 2014, p. 607).

Range cameras or color and range measuring RGB-D cameras have also been designed to overcome this limitation and to measure the range to the targets. However, most of these sensors are designed solely for indoor use and possess limited utility in outdoor lighting conditions, such as incident sunlight (Halmetschlager-Funek, Suchi, Kampel, & Vincze, 2018; Fu et al., 2020).

Because of range measurement limitations and challenges in difficult weather and lighting conditions, maximum potential can be achieved by fusing the data they provide with the data from radar or lidar systems (Campbell et al., 2018).

Radar

Radars are mostly used for remote sensing from air or space (see, e.g., Sinha, Jeganathan, Sharma, & Nathawat, 2015; J. Hyyppä et al., 2016). However, vehicle-mounted radar technology has been developed for autonomous driving purposes (see, e.g., Patole, Torlak, Wang, & Ali, 2017). Proposed applications for radar as an environmental sensor began with predictive crash sensing, obstacle detection, and braking, and continued to more complex functions, such as autonomous driving. The advantages of radar include its ability to measure range and velocity directly and sense long distances ahead, and its robustness to bad weather and poor lighting conditions (Hakobyan & Yang, 2019). An extra benefit of radar technology in a forest environment is its ability to sense through foliage (Hyyti, 2012; Duarte et al., 2022). Inside forests, radars have been used for obstacle detection (Hellström et al., 2009) and measurement of cutting height for standing trees (Rouveure, Faure, Marionneau, Rameau, & Moiroux-Arvis, 2014) and soil (Sucre, Tuttle, & Fox, 2011) and tree-root depth (Hruska, Čermák, & Šustek, 1999), to name but a few examples.

Unfortunately, the angular resolution of radars is usually much lower than that of optical sensors because the wavelength of the electromagnetic spectrum used in radars is longer. The most commonly used frequency bands for automotive radar are 24 and 77 GHz, of which most manufacturers are shifting toward the higher 77 GHz frequency (Hakobyan & Yang, 2019). This frequency allows for better angular resolution with smaller antenna, but is still limited to a few degrees. For example, Hasch et al. (2012) has shown that a 50×50 mm antenna aperture results in an angular separability of 17.5° at 24 GHz and 5.4° at 77 GHz.

It is also possible to attempt to combine the radar image formation problem with position estimation using SLAM methods, as shown by Rouveure, Monod, and Faure (2008) and Vivet et al. (2013). However, radar data are extremely noisy, and ground surface reflections cause undesired background signals (i.e., ground clutter), which are difficult to estimate from the data alone (Vivet et al., 2013). Therefore, in SLAM, both landmark extraction and data association can be false, causing the method to be unreliable. For example, the speckle effect, which is a randomness in the intensity caused by the mutual interference of a set of coherent wavefronts (Argenti, Lapini, Bianchi, & Alparone, 2013), can lead to false detections or false disappearances due to the different possible combinations of the radar signals measured (Vivet et al., 2013).

Sonar

In addition to radio waves, sound waves can also penetrate leaves and needles in the forest (Hyyti, 2012). When sound waves are used for echolo-

cation, sonars sensors can be utilized in a roughly similar manner to radars with radio waves. In robotics, low-cost ultrasonic rangefinders have been used to measure the distance to a nearby object by emitting a sound wave at an ultrasonic frequency and listening for that sound wave to bounce back (Mihelj et al., 2019, p. 102). Similar to other time-of-flight sensors, the elapsed time between the sound wave being generated and the sound wave bouncing back is used to calculate the distance between the sensor and the object. Sound possesses a rather slow velocity in air (approximately 343 m/s), so echo location based on sound is itself slow, and the method is only applicable for short distances. In addition, ultrasonic signals attenuate significantly in air, which makes accurate long-range (> 10 m) sensing difficult with sonar sensors (Kerstens, Laurijssen, & Steckel, 2019). Sonars are primarily used for underwater surveying, where sound travels faster (see, e.g., Hayes & Gough, 2009; R. E. Hansen, 2013). Only a few researchers have attempted to use synthetic-aperture sonar (SAS) in air (e.g., Saruwatari & Komura, 1999; Kerstens et al., 2019). In this work, ultrasonic sensors are found usable only at very short distances. Specifically, an ultrasonic rangefinder has been used in a safety mechanism for a camera lift in a point-cleaning tool (see Section 4.2).

Lidar

Lidars are among the most promising remote sensing equipment for the forest environment, as they can provide dense and accurate 3D measurements. A lidar measures distances on the principle of time of flight (ToF) by emitting a laser light and measuring the time it takes for the signal to reflect back (Campbell et al., 2018). According to Behroozpour, Sandborn, Wu, and Boser (2017) and Royo and Ballesta-Garcia (2019), there are three common approaches to achieve this. The first and most straightforward way is to emit a short laser pulse towards the target and compute the distance from the roundtrip time of the pulse's echo. The second approach, which is also commonly used, is based on amplitude modulation of a continuous wave, in which the phase shift between the emitted and backscattered waves is measured to compute the distance to the target. The third method is a frequency-modulated continuous-wave technique that is more commonly used in radar and sonar sensors. In this technique, the frequency of the laser light is modulated, and an interferometric detection scheme is used in the receivers (Uttam & Culshaw, 1985).

From the previous three methods, pulsed lidar can provide centimeter-level resolution for each pulse from short to long ranges, as nanosecond-long pulses often provide high instantaneous peak power (Royo & Ballesta-Garcia, 2019). The continuous wave method, on the other hand, uses longer amplitude modulated pulses, which limits the range separability of multiple successive reflections. This can be a problem in forests with abundant

underbrush and foliage which occlude the ground surface, tree stems, and branches. The last method is technologically the most complex to use, but it presents two outstanding benefits over the other techniques: firstly, it achieves better resolutions (down to 150 µm) even at long distances, and, secondly, it is able to obtain direct velocity measurements simultaneously for range data using the Doppler effect (Behroozpour et al., 2017; Royo & Ballesta-Garcia, 2019).

Most low-cost lidar sensors are pulsed sensors that can achieve a range accuracy of a few centimeters (e.g., ±3 cm for Velodyne Lidar, 2019b; Hesai, 2020; Ouster, 2021). This is a benefit, especially in a forest environment, since the pulse-type technique offers good range separability of multiple successive reflections if sufficiently short pulses are used. These are currently the most promising sensors for forest machine use. However, the lidar technology develops nowadays with huge steps since global investors have invested significantly to support the development of self-driving vehicles (Yeong et al., 2021). A comprehensive review by Holzhüter, Bödewadt, Bayesteh, Aschinger, and Blume (2023) lists various modern technologies for automotive lidar sensors which might be usable in forest machinery in the near future.

Lidar measurements can be used to generate 3D representations of the surrounding environment by registering multiple measurements in a point cloud when the position and the orientation of the laser emitter/receiver is known. Modern lidar sensors are capable of measuring distances at rates greater than one megahertz over longer ranges of over 100 meters (e.g., Velodyne Lidar, 2019a; Campos et al., 2020). Compared to the alternative range-measuring techniques previously introduced in this chapter, such as radars and sonars, lidars possess superior angular resolution. The beam divergence, which relates to angular separability, can be a fraction of a milliradian in state-of-the-art lidar sensors (Campos et al., 2020) and a few milliradians in low-cost sensors (e.g., 3×1.5 mrad for Velodyne Lidar, 2018). Lidar-based mapping is also more reliable and accurate than radar-based mapping (see, e.g., Hillier, Ryde, & Widzyk-Capehart, 2015) because multiple reflections or the speckle effect do not exert such a large effect on optical frequencies (Vivet et al., 2013).

As mentioned in the background section, lidars have often been utilized for the remote sensing of forests. They are commonly used from airplanes, in which context the process is called airborne laser scanning (ALS). In ALS, tightly integrated GNSS and inertial navigation sensors are used to estimate the position and attitude of the aircraft when a flight-management system helps the pilot fly a predefined path (Vosselman & Maas, 2010, pp. 22–23). Flying altitude is usually between 200 m and 4000 m, and the measurement density at ground level is usually from less than 1 to about 30 measurements per square meter (Vosselman & Maas, 2010, pp. 35–37).

In ALS lidar sensors, beam divergences are typically between 0.1 mrad and 1 mrad (Vosselman & Maas, 2010, p. 26). For example, if the divergence is 0.2 mrad, the footprint of the laser pulse will be 0.2 m from a survey height of 1000 m. For elevation data collection, the standard accuracy (in the local coordinate system) is 0.05–0.20 m for height and 0.2–1.0 m for position (Vosselman & Maas, 2010, p. 35). In ALS, lidar sensors are usually pulse-type ToF instruments (Vosselman & Maas, 2010, p. 37).

When a laser scanner is used in a stationary setting on the ground, the approach is known as terrestrial laser scanning (TLS). In contrast to ALS, which requires only one scanning direction (the other being accomplished by the moving aircraft), a TLS sensor must scan both horizontally and vertically (Vosselman & Maas, 2010, p. 37). In TLS, the sensor is usually used in a fixed position (e.g., on a tripod) and accurately positioned with the traversing and resection methods used as standard practice by land surveyors (C. K. Toth & Petrie, 2018). In addition, a total station and a GNSS receiver can be used to measure the sensor location. In a typical TLS instrument, the scanner measures the surrounding environment stepwise using rapid vertical mirror rotation and slower horizontal rotation to cover a full 360° field of view (FoV) (X. Liang et al., 2016). Multiple 3D point clouds are collected around the same forest plot and then registered to each other using either fixed-point locations, for example, special markers or reflectors added to the environment before scanning (Vosselman & Maas, 2010, p. 111), or a data-based registration procedure (e.g., Dold & Brenner, 2006; Barnea & Filin, 2008).

TLS is able to provide the best quality terrestrial point clouds for tree attribute estimation (X. Liang et al., 2018). From this data, forest inventory information (i.e., stand attributes) can be extracted with acceptable accuracies (e.g., tree diameter at breast height (DBH) and stem curve can be measured with root mean square errors (RMSE) of 1–2 cm) (X. Liang et al., 2016, 2018; J. Hyyppä et al., 2018). In addition, may other attributes, including forest density, volume, upper stem diameters, height of crown base, basal area, and biomass, can be estimated (Watt & Donoghue, 2005; J. Hyyppä et al., 2018). However, tree height estimation is one of the most difficult parameters to estimate using TLS (X. Liang et al., 2018). Forest inventory can be collected either using an area-based method, where average values for each plot are calculated, or at the single-tree level, where each tree is measured and mapped separately (Kankare et al., 2017).

TLS is an accurate but slow and expensive method for remote sensing, since it requires the static position of the sensor to be known when measuring. Therefore, it is not directly usable for forest machine automation. A faster but more difficult way to produce a similar 3D point cloud is to move the sensor while collecting measurements and continuously measure the location and orientation of the sensor (Kukko et al., 2007). This is called mobile laser scanning (MLS) (Pu, Rutzinger, Vosselman, & Elberink, 2011;

Kukko, Kaartinen, Hyyppä, & Chen, 2012), but the terms personal laser scanning (PLS) (X. Liang et al., 2014) and kinematic laser scanning (KLS) (Kukko, Kaartinen, & Hyyppä, 2020) have also been used. The difficulty of this method stems from the sufficiently accurate estimation of the position and orientation of the sensor during data collection. In open air, accurate dynamic positioning is simpler, and it is already used in, for example, ALS (Vosselman & Maas, 2010, p. 22). However, as already noted in Section 2.2, position estimation under forest foliage is more difficult, and therefore SLAM-based approaches have also been extensively used in mobile and personal forest mapping in addition to GNSS and IMU solutions (Kukko et al., 2017; J. Hyyppä et al., 2018).

In addition to high precision individual tree-level measurements using TLS scanners (e.g., X. Liang et al., 2016), lidars have been used inside the forest for wheel rut measurements (Salmivaara et al., 2018), forest navigation (e.g., Ringdahl, 2007; Wooden et al., 2010), tree detection around the harvester (Sihvo et al., 2018), tree diameter measurement at breast height (Jutila, Kannas, & Visala, 2007; Zheng et al., 2012; Ringdahl, Hohnloser, Hellström, Holmgren, & Lindroos, 2013), stem mapping with a horizontal 2D laser scanner (Miettinen et al., 2007; Öhman et al., 2008; Rossmann et al., 2009; J. Tang et al., 2015), and increasingly also for the 3D mapping of forests (Kukko et al., 2017; Pierzchała et al., 2018; S. W. Chen et al., 2020). As a comparison, early attempts at mapping forests and measuring tree stems in 3D using a rotating laser scanner were already made by the author more than a decade ago (Hyyti, 2009) and also later in Publication VII.

3. Proprioceptive perception

Proprioceptive perception enables the robot to perceive its own state in the world. This chapter presents the developed proprioceptive perception methods and their sensor setups. The sensor setups include minimal instrumentation for crane posture estimation used in Publication I and dual-IMU instrumentation of a forest machine tool used in Publication III. Here, contrary to the scientific articles published by the author, the practical hardware innovations and industrial applicability of the sensor setups are elaborated. After the sensor setups, the proprioceptive perception methods are proposed. These include a particle filter for crane posture estimation in Publication I, and three extended Kalman filters, one for robust and adaptive attitude estimation using an IMU in Publication II, another for tool swaying angle estimation in Publication III, and the third for operator's head pose estimation in a forest machine cabin in Publication IV.

3.1 Minimal Instrumentation for Crane Posture Estimation

Publication I proposes novel instrumentation using a 2D laser scanner mounted vertically on the side of a forestry crane and a rotation encoder for the first joint angle (Slew in Figure 3.1). The setup enables direct observation of the crane tip position, from which the bending of the flexible crane can be estimated. Traditional robotic instrumentation using, for instance, electric motors and angular position sensors determines the tip position with an assumption of rigid links and measured joint angles between them. With a flexible crane, this causes large position offsets in the estimated tip position, as previously noted in Section 2.3. Additionally, the proposed instrumentation enables 3D scanning of the surrounding environment as the crane is turned from side to side.

Crane posture estimation is achieved with minimal occlusion to the environment by using two small round metal tubes as targets on the side of the forestry crane boom, as shown in Figure 3.1. The targets are magnetically attached onto the ends of steel pivot pins that work as joint

axles between the boom parts. The magnetically attached targets are designed to be easily installed and to drop off in case of direct contact with trees or other obstacles that would otherwise damage the targets or the pivot pins.

(a) 2D laser scanner

(b) Target 1

(c) Target 2

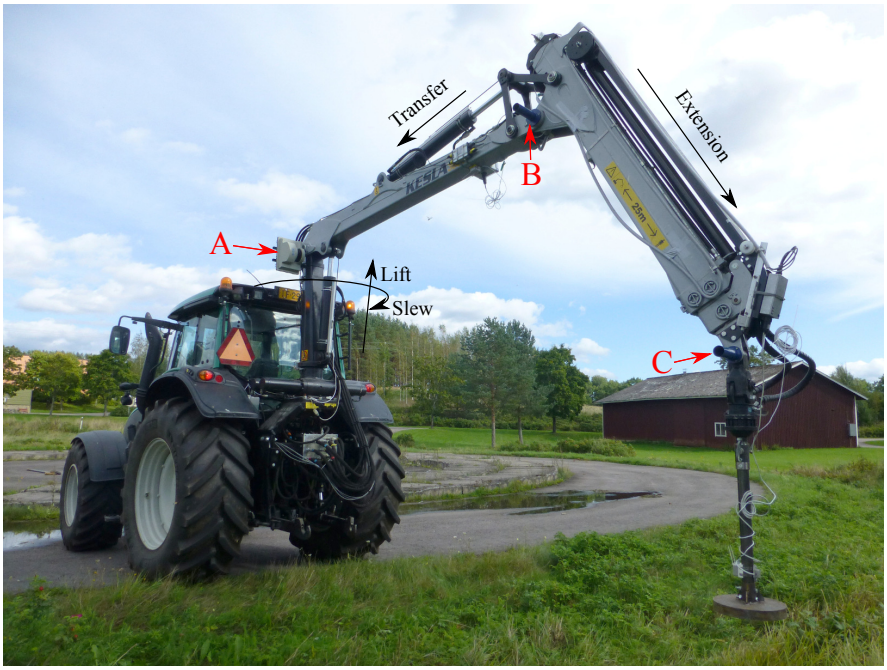


Figure 3.1. Overview of the system setup: A) 2D laser scanner, B) Target 1, C) Target 2. The names of joint angles and their positive directions are also shown.

A single target mounted on the side of the boom tip location (C in Figure 3.1) would reveal the boom tip location (e.g., the distance and angle to the lidar), but it is insufficient for defining the full posture of the crane boom with two rotating and one translating joints (Lift, Transfer, and Extension in Figure 3.1, respectively). Therefore, another target is added to the middle of the crane (B in Figure 3.1). Together, these two targets have the smallest possible occlusion to the surrounding environment, but, at the same time, they are able to reveal the full configuration of the crane joints in a single laser scan profile, as later shown in Figure 4.1. A third, redundant, target could be added to the side of the crane to increase the robustness of crane tracking to, for instance, the presence of occlusions. However, this would introduce ambiguity in defining each target, espe-

cially if some of the targets were occluded. Moreover, this would add an unnecessary obstacle to measuring the environment beyond the crane.

The targets and the environment beyond them are measured with a 2D laser scanner (LMS221 by SICK AG, 2006) (A in Figure 3.1). To provide free FoV towards the environment, the scanner is located at the highest point on the crane at which collisions with trees or obstacles remain unlikely. For example, an alternative position nearer the boom tip would allow closer and higher observation of the targets and environment, but there the scanner would be vulnerable to impacts with branches during forest use. The laser scanner is configured in interpolated four scan mode to combine four adjacent scans at 1/4° intervals to provide a maximum angular resolution of 721 range measurements per its 180° FoV (SICK AG, 2006). Any similar laser scanner could be used instead, if it provides at least the same angular resolution of 0.25° on at least the same FoV, an equivalent range measuring accuracy, and if it tolerates the environmental and vibration conditions on the forest machine.

The diameter of the two similar painted-metal tubes working as targets 1 and 2 (B and C in Figure 3.1) is 60 mm. The tube design is a compromise suited for the selected sensor. On one hand, it is as small as possible, to occlude minimally, but, on the other hand, it is sufficiently large to be clearly detected and measured from the longest crane reach of approximately 8 meters from the scanner. The paintwork of the targets is designed to avoid overly bright reflection that could dazzle the scanner (i.e., a Lambertian surface is preferred over retroreflective material). This also minimizes any effect on the range measurements from the environment behind, nearby, or partly occluded by the targets. Both targets are defined similar in shape and paintwork to simplify the detection and measurement of the targets in the method explained later in Section 3.2. This choice enables the method to use the same measurement process for both targets simultaneously, but the drawback is that different colors, sizes, nor shapes cannot be used to distinguish between the targets. The target positioning accuracy of the laser scanner is also checked in Publication I for these targets to verify that the target does not cause any bias to the position estimates. Although the target measuring method does not need any calibration, the position and orientation of the scanner need to be known precisely or calibrated using, for example, the data-driven method explained in Publication I.

The proposed sensor setup may also be applicable in any equivalent crane which has the boom in one line allowing a free optical path next to the crane in all crane postures. Depending on the applied crane, the target positions as well as the model parameters would need to be redefined in the kinematic model used in the crane posture estimation method defined later in Section 3.2.

3.2 Particle Filter for Crane Posture Estimation

To robustly estimate the posture of a flexible hydraulic crane (see Section 3.1), an SIR-type particle filter (see Nonlinear Filtering under Section 2.1 for details) was developed in Publication I with custom measurement and system equations. A SIR-type particle filter is used since it is simple to implement and it can handle the multi-modal non-Gaussian probability estimates present in the system as found earlier in Section 2.1. The proposed algorithm is called a *Crane Posture Particle Filter* (CPPF).

The state $\mathbf{x}$ estimated in CPPF is

$$\mathbf{x} = [\theta_2 \;\; \theta_3 \;\; d_4]^\mathsf{T}, \tag{3.1}$$

where θ_2 and θ_3 are the two joint angles in the crane boom and d_4 is the extension length. In the setup, the first joint angle of the crane (θ_1) is measured directly using an angle position sensor (see Section 3.1).

The control input $\mathbf{u}$ for the CPPF equals a joint velocity vector,

$$\mathbf{u} = \hat{\mathbf{v}}_{joint} = \left[\dot{\theta}_2 \;\; \dot{\theta}_3 \;\; \dot{d}_4\right]^\mathsf{T}, \tag{3.2}$$

where $\dot{\theta}_2$ and $\dot{\theta}_3$ are angular velocities of the joint angles and $\dot{d}_4$ is a linear velocity of the extension joint. The joint velocities ($\hat{\mathbf{v}}_{joint}$) are not directly measured but instead estimated using control signals sent to the hydraulic valves and a nonlinear model to transform those control signals into predicted cylinder velocities, which are then transformed into joint velocities using an inverse kinematic model of the crane (see Publication I for the specific models of the crane used). Alternatively, these joint velocities could also be estimated with inertial measurements, as in Publication III.

The measurement of the CPPF is

$$\mathbf{y} = \left[r_1, r_2, ..., r_{N_l}\right]^\mathsf{T}, \tag{3.3}$$

where r_l, $l \in \{1, 2, ..., N_l\}$, are the N_l range measurements provided by the laser scanner from its FoV during one scan (see Section 3.1).

After initialization, the CPPF algorithm consists of four repeated phases: prediction, update, normalization, and resampling:

0) **Initialization** sets particles uniformly distributed around the state space. All the weights w_i are set to an equal weight of $1/N_p$, where N_p is the number of particles in the filter. In addition, one particle is changed to match each of the crane posture hypotheses calculated from target candidates found with a separate, deterministic target detector method to boost the initialization phase such that one particle is placed around each posture hypothesis. The deterministic target detector fits known size circles to the lidar scan and uses kinematic constraints of the crane to find a suitable crane posture (see Publication I for details).

1) **Prediction** is performed using a dynamic model of the crane and the control signals of the hydraulic valve to predict crane posture change from a previous time to the present moment. Since prediction is prone to error, noise is added to the system for each $i = 1,\ldots,N_p$ by drawing predicted particles $\mathbf{x}_i^-(k)$ from a normal distribution,

$$\mathbf{x}_i^-(k) \sim \mathcal{N}\left(\mathbf{x}_i(k-1) + \mathbf{1}_i \hat{\mathbf{v}}_{joint}(k)\Delta t(k),\ \mathbf{\Sigma}\right), \tag{3.4a}$$

where

$$\mathbf{1}_i = \mathrm{diag}\left(I_{i,1}, I_{i,2}, I_{i,3}\right),$$

$$I_{i,j} = \begin{cases} 1, & U < 0.9, \quad U \sim \mathcal{U}(0,1) \\ 0, & \text{otherwise} \end{cases} \qquad \forall j \in \{1,2,3\}. \tag{3.4b}$$

In Equation (3.4a), $\mathbf{x}_i(k)$ is the ith particle in a three-dimensional state space $\mathbf{S}$ defined in Equation (3.1) at time step k. In the equation, the mean of the normal distribution is located at the previous state $\mathbf{x}_i(k-1)$, which is updated by adding the movement predicted with the joint velocity $\hat{\mathbf{v}}_{joint}(k)$ during a sample period $\Delta t(k)$ (a period between two adjacent indices $k-1$ and k).

In Equation (3.4), $\mathbf{1}$ is a random indicator function to reduce the velocity measurement to zero for 10% of randomly selected particles for each state independently (indicated with an index j). $\mathbf{1}$ is implemented in Equation (3.4b) by sampling a standard uniform distribution ($\mathcal{U}(0,1)$) for each i and j separately. This modifies the distribution such that there are two independent modes for each state: the first (90% of the probability mass) at the predicted position, and the other (10%) at the previous position with the same standard deviation. The ratio of 10% was found to function best in practice in the tested system. This modification is required to handle occasions when control signals indicate a movement, but the crane does not move accordingly. These kinds of events may be caused, for example, by joint limits or by obstacles that restrict the movement of the crane.

As, for simplicity, the states are assumed to be independent of each other, the covariance matrix,

$$\mathbf{\Sigma} = \left(\Delta t(k)\right)^2 \mathrm{diag}\left(\sigma_{\dot{\theta}_2}^2, \sigma_{\dot{\theta}_3}^2, \sigma_{\dot{d}_4}^2\right), \tag{3.5}$$

is defined to contain variances on the main diagonal and zeros elsewhere. These variances are constructed from the sample period $\Delta t(k)$ and standard deviations of velocity errors for each state: $\sigma_{\dot{\theta}_2}$, $\sigma_{\dot{\theta}_3}$, $\sigma_{\dot{d}_4}$. These variances were estimated from data in Publication I.

At the end of each prediction step, invalid state configurations are detected, and their corresponding weights are set to zero. For this purpose, a subset of feasible states $\mathbf{S_f} \subset \mathbf{S}$ is defined, where $\theta_2 \in [\theta_{2,min}, \theta_{2,max}]$,

$\theta_3 \in [\theta_{3,min}, \theta_{3,max}]$, and $d_4 \in [d_{4,min}, d_{4,max}]$ (see Publication I for the values used). Then a predicted weight $w_i^-(k)$ is used to zero the effect of infeasibly located particles using the following equation:

$$w_i^-(k) = \begin{cases} w_i(k-1), & \mathbf{x}_i^-(k) \in \mathbf{S_f} \\ 0, & \text{otherwise.} \end{cases} \tag{3.6}$$

2) The **update** step uses measurements $\mathbf{y}(k) \in \mathbb{R}^{N_l}$ conditioned on the predicted states $\mathbf{x}_i^-(k) \in \mathbb{R}^3$ to compute weights $w_i(k) \in \mathbb{R}$. These are obtained from measurement likelihoods as

$$w_i(k) = w_i^-(k)\mathscr{L}\left(\mathbf{y}(k)|\mathbf{x}_i^-(k)\right). \tag{3.7}$$

The measurement likelihood $\mathscr{L}\left(\mathbf{y}|\mathbf{x}_i^-\right)$ is derived in Publication I to be proportional to an inverse measurement model $P\left(\mathbf{x}_i^-|\mathbf{y}\right)$, which can be approximated in the proposed case of two targets as

$$P\left(\mathbf{x}_i^-|\mathbf{y}\right) \approx \prod_{\forall j \in \mathbf{T}} \sum_{l=1}^{N_l} \delta_{l,i,j} L\left(r_{i,j}|r_l\right), \tag{3.8a}$$

where

$$\delta_{l,i,j} = \begin{cases} 1, & \text{when } |\phi_l - \phi_{i,j}| < \Delta\beta/2 \\ 0, & \text{otherwise.} \end{cases} \tag{3.8b}$$

In Equation (3.8), $\delta_{l,i,j}$ is defined as equal to 1 only when the angles of a laser range observation (ϕ_l) and an expected target ($\phi_{i,j}$) match within the laser scanner resolution ($\Delta\beta$). For a laser scan, this occurs for a single configuration of indexes l and i, for both targets j if the targets are in the field of view of the scanner. Thus, only one matching range measurement is compared with each target $j \in \mathbf{T}$ using a one-dimensional fitness function $L\left(r_{i,j}|r_l\right)$, which is an approximation of the likelihood of a range of an expected target $r_{i,j}$ given the matched laser range observation r_l (including the precomputed measures N_l and C_l defined later). It is defined as

$$L\left(r_{i,j}|r_l\right) = \begin{cases} r_{i,j}\, l_{miss}, & d_{i,j,l} > \epsilon_d \\ l_{step} + \left(1 - l_{step}\right) L_{target}\left(r_{i,j}, r_l, N_l, C_l\right), & d_{i,j,l} \le \epsilon_d \end{cases}, \tag{3.9a}$$

where

$$d_{i,j,l} = r_l + a - r_{i,j}. \tag{3.9b}$$

In Equation (3.9), r_l is the laser range measurement, $r_{i,j}$ is the range of the expected target, and $d_{i,j,l}$ is the distance between the expected target and the target indicated by the range measurement ($r_l + a$). The fitness function is constructed from a step function which increases the likelihood of an expected target to l_{step} after the target can be fitted

behind the range measurement (with a safe margin ϵ_d). If the target is seen through (i.e., the range indicated by the measurement is larger than the expected range), the likelihood is defined to linearly increase as a function of the expected range with a constant likelihood of missing the target l_{miss}. In Publication I, l_{step} value of 0.5 and l_{miss} value of 0.01 were used.

In Equation (3.9a), L_{target} is a target-likeness measure, which is a function of the expected range $r_{i,j}$, the observed range r_l, the precomputed cluster-size measure N_l, and the precomputed count-to-the-middle measure C_l. It is constructed as a product of three exponential functions:

$$L_{target}\left(r_{i,j}, r_l, N_l, C_l\right) = L_{dist}\left(d_{i,j,l}\right) L_{size}\left(N_l, r_{i,j}\right) L_{mid}\left(C_l, r_{i,j}\right), \quad (3.10a)$$

where

$$L_{dist}\left(d_{i,j,l}\right) = \exp\left(-d_{i,j,l}^2 / \sigma_{dist}^2\right), \quad (3.10b)$$

$$L_{size}\left(N_l, r_{i,j}\right) = \exp\left(-\left(N_l \Delta\beta r_{i,j}\right)^2 / \sigma_{size}^2\right), \quad (3.10c)$$

$$L_{mid}\left(C_l, r_{i,j}\right) = \exp\left(-\left(C_l \Delta\beta r_{i,j}\right)^2 / \sigma_{mid}^2\right). \quad (3.10d)$$

In Equation (3.10), $d_{i,j,l}$ is the difference defined in Equation (3.9b), $r_{i,j}$ is the range of the expected target, $\Delta\beta$ is the angular resolution of the laser scanner, N_l is the precomputed cluster-size measure, in detail, the amount of middle points associated with a cluster in a separate, deterministic target detector method (see Publication I for details), and C_l is the precomputed count-to-the-middle measure. It is computed as a count from the current index l to the center of the current cluster of middle points.

The first exponential function (L_{dist}) in Equation (3.10b) weights the right distance behind the range measurement to fit a target. Its parameter σ_{dist} should be near the radius of the expected target (50 mm in Publication I). The second function (L_{size}) in Equation (3.10c) weights small clusters and discards overly large clusters. Its parameter σ_{size} should be tuned to be sufficiently large to avoid filtering target-size clusters (400 mm in Publication I). The last exponential function (L_{mid}) in Equation (3.10d) then weights the locations at the centers of clusters. Its parameter σ_{mid} should also be near the radius of the target (30 mm in Publication I). This last function is a heuristic which is added to keep the area of the likely locations of targets similar in small and large clusters of range observations. It is based on the assumption that, in target sized clusters, the target is most likely to be located at the center. When all these exponential functions have their maximal value, the fitness function $L\left(r_{i,j}|r_l\right)$ in Equation (3.9) produces the best fit. Note that each of these exponential functions can have values between 0 and 1.

3) **Normalization** divides unnormalized weights $w_i(k)$ by the sum of all updated weights to obtain the normalized weights for each i:

$$W_i(k) = \frac{w_i(k)}{\sum_{i=1}^{N_p} w_i(k)}. \tag{3.11}$$

4) **Resampling** generates a new set of particles $\mathbf{x}_i(k)$ by drawing them among the predicted particles $\mathbf{x}_i^-(k)$ according to the normalized weights $W_i(k)$. This can be efficiently achieved using an inverse transform sampling method (Arulampalam et al., 2002). After resampling, all weights $w_i(k)$ are set to an equal value of $1/N_p$.

After the fourth phase, the first phase is re-entered for the next time step. The best crane posture estimate is computed from the set of all particles after the normalization phase using a kernel density estimate with Gaussian kernels (Musso, Oudjane, & Le Gland, 2001). More specifically, the kernel density estimate $D_i(k)$ is computed for each particle at time step k using an equation

$$D_i(k) = \sum_{j=1}^{N_p} W_j(k) \exp\left(\frac{-\|\mathbf{x}_i(k) - \mathbf{x}_j(k)\|^2}{\sigma_{kde}^2}\right), \tag{3.12}$$

where $W_i(k)$ is a normalized weight and $\mathbf{x}_i(k)$ is the corresponding ith particle. In the equation, σ_{kde}^2 is the variance of the kernels in the density estimate. The σ_{kde} value used was 0.03, which is about 1.7° for the first two angular states and 3 cm for the extension. For simplicity, the same kernel size parameter was used for each of the three states, as the probability distributions were roughly similar in each dimension. Finally, a maximum a posteriori (MAP) estimate is used to select the particle with the largest kernel density estimate, $D_i(k)$ in Equation (3.12). The state represented by this particle is taken to be the best crane posture estimate.

The value of the MAP estimate,

$$\hat{D}(k) = \max_{i=1\ldots N_p} (D_i(k)), \tag{3.13}$$

is limited between zero and one, and it approaches one when all particles are at the same place. Conversely, when all particles are dispersed, the value approaches zero. The benefit of the MAP estimate value is two-fold. In a relative sense, the largest value gives the particle closest to the center of the cluster yielding the best estimate for the target position. In an absolute sense, it acts as a *quality self-measure* to distinguish between unreliable ($\hat{D} \sim 0$) and reliable ($\hat{D} \sim 1$) detections. If this value drops below the threshold $C_{reinit} = 0.35$, a re-initialization is performed. The parameter value of 0.35 was tuned manually on the tested setup to handle all faults but not to cause unnecessary re-initialization. This reinitialization is similar to the Initialization step, except that the particles are not redistributed

uniformly. The previous low-quality estimate is simply enhanced with the crane posture hypotheses calculated using target candidates found with a separate target detector method (see Publication I for details).

3.3 Robust and Adaptive Attitude Estimation for an IMU

In the DCM IMU algorithm proposed in Publication II, the rotation from the body-fixed frame to the navigation frame is represented as a direction cosine matrix (DCM) of Euler angles according to the Tait-Bryan ZYX convention,

$$ {}^n_b\mathbf{C} = \begin{bmatrix} \theta_c\psi_c & -\phi_c\psi_s + \phi_s\theta_s\psi_c & \phi_s\psi_s + \phi_c\theta_s\psi_c \\ \theta_c\psi_s & \phi_c\psi_c + \phi_s\theta_s\psi_s & -\phi_s\psi_c + \phi_c\theta_s\psi_s \\ -\theta_s & \phi_s\theta_c & \phi_c\theta_c \end{bmatrix}, \tag{3.14} $$

where subscript s denotes the $\sin(\cdot)$ function and c denotes the $\cos(\cdot)$ function. The angles on which the functions operate are roll ϕ, pitch θ, and yaw ψ, and they are defined as angles around the x, y, and z axes in the body-fixed frame, respectively. This rotation formalism is selected because the lowest row of ${}^n_b\mathbf{C}$ indicates the direction of gravity in the body-fixed frame, as the navigation frame is defined to lie such that z is pointing upwards.

In the proposed EKF, the direction of gravity (as the lowest row of ${}^n_b\mathbf{C}$) and gyroscope biases (i.e., angular velocity offsets) b_x, b_y, and b_z (around x, y, and z axes in the body-fixed frame, respectively) are included in the state vector

$$ \mathbf{x} = \begin{bmatrix} {}^n_bC_{31} & {}^n_bC_{32} & {}^n_bC_{33} & b_x & b_y & b_z \end{bmatrix}^\mathsf{T}. \tag{3.15} $$

For an IMU in a body-fixed frame, six measurements are available, namely accelerations a_x, a_y, and a_z in x, y, and z directions, respectively, and angular velocities ω_x, ω_y, and ω_z around the same axes. To minimize the length of the state vector to enable efficient computation, gyroscope-measured angular velocities are fed to the EKF as a control input vector,

$$ \mathbf{u} = \begin{bmatrix} \omega_x & \omega_y & \omega_z \end{bmatrix}^\mathsf{T}, \tag{3.16} $$

and the accelerometer measurements are used as a measurement vector

$$ \mathbf{y} = \begin{bmatrix} a_x & a_y & a_z \end{bmatrix}^\mathsf{T}. \tag{3.17} $$

In the prediction step of the proposed EKF, the following nonlinear system model (see the nonlinear filtering part of Section 2.1 for details) is used to predict the state from the previous time step to the current one:

$$ f_k(\mathbf{x}_k, \mathbf{u}_k) = \begin{bmatrix} \mathbf{I}_3 & -[{}^n_b\mathbf{C}_3\times]_k\Delta t \\ \mathbf{0}_{3\times3} & \mathbf{I}_3 \end{bmatrix} \mathbf{x}_k + \begin{bmatrix} [{}^n_b\mathbf{C}_3\times]_k\Delta t \\ \mathbf{0}_{3\times3} \end{bmatrix} \mathbf{u}_k, \tag{3.18} $$

where Δt is the sampling interval and $[{}^n_b\mathbf{C}_3\times]_k$ is a skew-symmetric matrix generated from the first three state variables at time index k:

$$[{}^n_b\mathbf{C}_3\times]_k = \begin{bmatrix} 0 & -{}^n_bC_{33} & {}^n_bC_{32} \\ {}^n_bC_{33} & 0 & -{}^n_bC_{31} \\ -{}^n_bC_{32} & {}^n_bC_{31} & 0 \end{bmatrix}_k . \tag{3.19}$$

Since the state is designed to indicate the direction of gravity, and since the largest share of the acceleration measurement is assumed to be caused by the Earth's gravity, the measurement model of the proposed EKF is simple, and it can be represented with a linear model using a matrix (see Kalman Filtering in Section 2.1 for details):

$$\mathbf{H} = \begin{bmatrix} g\mathbf{I}_3 & \mathbf{0}_{3\times 3} \end{bmatrix}, \tag{3.20}$$

which is employed to predict the triaxial accelerometer measurement in Equation (3.17) using the predicted state. In Equation (3.20), g is the magnitude of the Earth's gravity.

As the measurement model of the proposed EKF is only valid at rest (with no other forces than gravity present), the algorithm produces biased estimates when non-gravitational acceleration is present in the accelerometer measurements. As the method is designed to be used in motion, in Publication II, the limitation is relaxed by estimating the share of non-gravitational acceleration and adapting the measurement covariance matrix $\mathbf{R}_k$ to take account of the credibility of the measurement. In the proposed EKF, the covariance matrix is adapted with

$$\mathbf{R}_k = \left(\|\mathbf{a}_k\|\sigma_a^2 + \sigma_f^2\right)\mathbf{I}_3, \tag{3.21}$$

where $\|\mathbf{a}_k\|$ is the magnitude of the estimated non-gravitational acceleration, σ_a^2 is a variance added to account for non-gravitational acceleration, and σ_f^2 is the variance of the accelerometer measurement (see Publication II for parameter values). Non-gravitational acceleration is estimated as the difference between the current measurement ($\mathbf{y}_k$) and the estimate of the predicted measurement ($\hat{\mathbf{y}}_k^- = \mathbf{H}\hat{\mathbf{x}}_k^-$) using

$$\mathbf{a}_k = \mathbf{y}_k - \mathbf{H}\hat{\mathbf{x}}_k^-, \tag{3.22}$$

where $\hat{\mathbf{x}}_k^-$ is the predicted state before measurement update (see Kalman Filtering in Section 2.1 for details).

The state $\mathbf{x}$ includes the lowest row of the rotation matrix in Equation (3.14), which is defined as a unit vector. Since the EKF does not, by default, take this constraint into account, it must be added to the system. In Publication II, this is achieved by normalizing the length of the first three state variables and projecting this into the state covariance matrix. This process with all the required equations is demonstrated in the Appendix of Publication II.

Finally, the attitude in Euler angles can be computed from the state variables for roll ϕ and pitch θ using

$$\begin{aligned} \phi_k &= \operatorname{atan2}\left({}^n_b C_{32,k},\ {}^n_b C_{33,k}\right) \\ \theta_k &= \arcsin(-{}^n_b C_{31,k}), \end{aligned} \tag{3.23}$$

where atan2 is an inverse tangent function with two arguments to distinguish angles in all four quadrants (R. S. Jones, 1991).

The yaw angle is more difficult to estimate, since the initial heading of the IMU is unknown (initialized as zero in the EKF) and as gravity provides no information about the heading direction. As found in Section 2.2, many alternative attitude estimation algorithms use other sensors, such as magnetometers, to enable heading estimation. However, in forest machinery, which is commonly built from ferromagnetic metal, contains moving parts and changing electric current flows, and can operate under power lines, dependence on magnetometers is not a plausible solution. Therefore, here, the yaw angle is integrated from bias-corrected angular velocities. This integration can be achieved, for example, also by keeping the first row of the rotation matrix in memory and using the cross product to generate the missing middle row, as exemplified later in Section 3.6.

After the full rotation matrix ${}^n_b\mathbf{C}_k$ is generated, the yaw angle ψ can be calculated using

$$\psi_k = \operatorname{atan2}\left({}^n_b C_{21,k},\ {}^n_b C_{11,k}\right). \tag{3.24}$$

Note that this yaw angle is relative to the initial position and will most likely contain drift caused by the residual errors in the gyroscope measurements. However, as shown later in Section 3.6, other measurements can be combined with the estimate to overcome this limitation.

3.4 Dual IMU Instrumentation of a Forest Machine Tool

The attitude of an IMU can be estimated quite robustly with only one IMU, as shown in Publication II. However, as explained in Section 2.2, heading and also the bias around the gravity vector remain unobservable and thus unknown in an IMU, where only accelerometer measurements are used to align the sensor in the Earth's gravitational field. Since the tool freely hanging from the tip of the forestry crane can rotate in all three directions (swaying back and forth (α), side to side (β), and rotating around its upward-pointing axis (γ)), the pose of the tool is not fully observable using a single IMU attached to the tool (see Figure 3.2).

As also noted in Section 2.2, a magnetometer is commonly utilized in AHRS sensors to simplify gyroscope bias estimation and enable heading measurement by using the Earth's magnetic field. However, it cannot be used in this case. The forestry crane and the rotator link are constructed

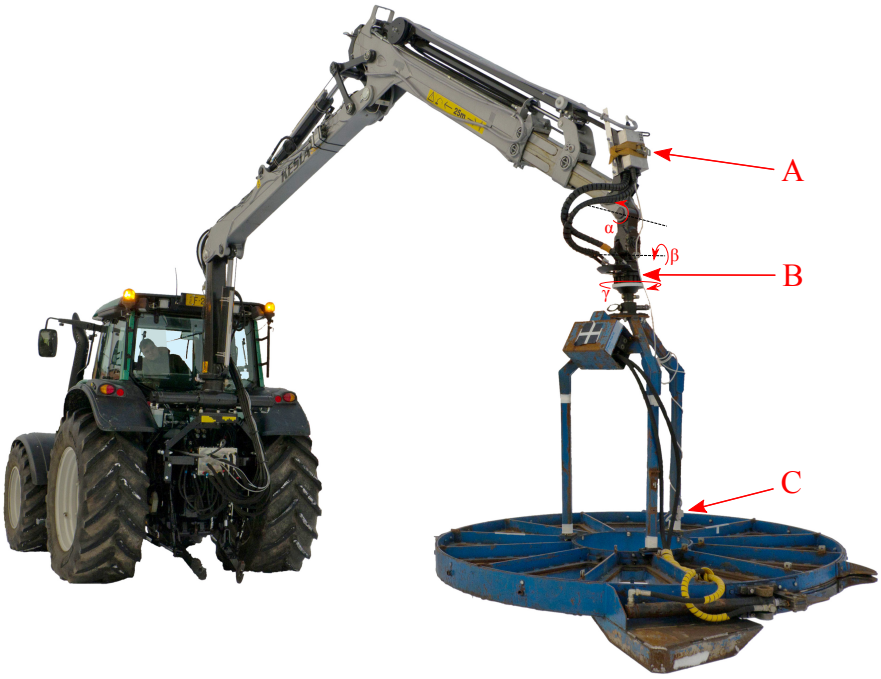


Figure 3.2. The forest machine with the cutting tool and the dual IMU instrumentation (A and C) to estimate the dynamic pose of the tool on the freely swaying rotator link (B). The rotation angles α, β, and γ are also marked in the figure.

from steel, which distorts magnetic fields, and high-current electronic signals are occasionally present in nearby cabling, which renders the magnetometer unreliable. Furthermore, installing traditional position sensors on each of these joints is challenging due to the mechanical structure of the rotator link. In addition, the tool is often in contact with obstacles in the forest, and thus the sensors should be well protected or built inside the mechanism. Similarly to Kesla 305T in Figure 3.2, which is a typical forestry crane designed for loading tree trunks, most other forest machines have similar boom with an extension and a rotator link.

As a solution to this problem, Publication III proposes combined instrumentation with two similar low-cost IMU sensors for estimating the tool orientation. This is achieved by using the dynamics of the tool as a swaying pendulum to estimate all three free rotation parameters reliably. The two IMUs must be placed at both ends of the joint chain to enable the estimation of the three unknown joint angles in between. In the proposed solution, identical IMUs built using ADXL345 three-axis digital accelerometer (Analog Devices, 2012) and ITG-3200 three-axis digital gyroscope (InvenSense, 2010) chips were connected through a fast mode I2C interface to a microcontroller synchronizing the data from both sensors. The first IMU is mounted on the tip of the crane boom (A in Figure 3.2), and the

other is mounted on the tool (C in Figure 3.2) after the rotator link and the rotator (B in Figure 3.2), allowing rotation in three directions.

Two similar low-cost IMUs are used, since they provide data with similar noises and measuring frequencies, and their measurements can be synchronized. The two IMUs, which both measure triaxial accelerations and angular velocities, provide all the required measurements for estimating the three unknown joint angles in between them. Adding more sensors would only complicate the system and increase its cost. The used sensors were low-cost cellphone-grade sensors of their time. Newer low-cost single chip IMU sensors such as MPU-9250 (InvenSense, 2016) or other more capable sensors should also work.

The dimensions of the links between joints and the length of the pendulum to the center of mass of the tool are required to derive the dynamic equations of the system. These are used in the proposed EKF (see Section 3.5) in Publication III to estimate the joint angles of the tool. The estimate is computed at 100 Hz frequency in an embedded computer located in the same box with the first IMU (A in Figure 3.2). The estimated joint angles and their velocities are transmitted to the forest machine via a CAN bus built in the crane according to ISO 11783 standards (also referred to as ISOBUS). See more details on our ISO 11783 compatible forestry crane from Kalmari et al. (2013).

3.5 Kalman Filter for Tool Swaying Angle Estimation

The joint angles of a freely hanging point cleaning tool (see Section 3.4 are estimated with an EKF (see nonlinear filtering in Section 2.1 for details) that takes triaxial accelerometer and gyroscope measurements from two time-synchronized IMUs. To enable implementation on an embedded computer, filter computation is optimized by using some of the measurements as control inputs in the filter instead of increasing the state vector size. The state

$$\mathbf{x} = \begin{bmatrix} \alpha & \beta & \gamma & \dot{\alpha} & \dot{\beta} \end{bmatrix}^{\mathsf{T}}, \tag{3.25}$$

is kept minimal, consisting of only the swaying angles of the rotator link α and β, the angle of rotator motor γ, and the angular velocities of the first two angles $\dot{\alpha}$ and $\dot{\beta}$ (see Figure 3.3 for details).

The control input vector $\mathbf{u}$ for the EKF is

$$\mathbf{u} = \begin{bmatrix} \underbrace{\ddot{x}_m \quad \ddot{y}_m \quad \ddot{z}_m \quad \dot{\psi} \quad \ddot{\psi}}_{\mathrm{IMU}_1} \quad \underbrace{\dot{\gamma}}_{\mathrm{IMU}_2} \end{bmatrix}^{\mathsf{T}}, \tag{3.26}$$

where $\ddot{x}_m$, $\ddot{y}_m$, and $\ddot{z}_m$ are the accelerations, $\dot{\psi}$ is the angular velocity (around the z_m axis), and $\ddot{\psi}$ is the equivalent angular acceleration. In addition, the angular velocity of the rotator motor $\dot{\gamma}$ is chosen as a control input for the EKF. The required (non-gravitational) accelerations ($\ddot{x}_m$, $\ddot{y}_m$,

and $\ddot{z}_m$) and angular velocity $\dot{\psi}$ can be extracted from the boom tip IMU (IMU1 in Figure 3.3) measurements using an attitude estimation algorithm (see Section 3.3 and Publication II). The angular acceleration $\ddot{\psi}$ can then be numerically differentiated from the angular velocity estimate $\dot{\psi}$. The angular velocity of the rotator motor $\dot{\gamma}$ can be directly measured with the tool IMU (IMU2 in Figure 3.3) as angular velocity around the z axis, ω_z.

In Publication III, the system model is derived using a dynamic model of the swaying tool. A viscous friction model is used, where the torque is set proportional to the angular velocity. This model does not accurately represent real friction, but is simple to use and requires identification of only one friction parameter per joint. As the mass of the tool is assumed to be constant, it is separated from the friction parameters b_α and b_β. The

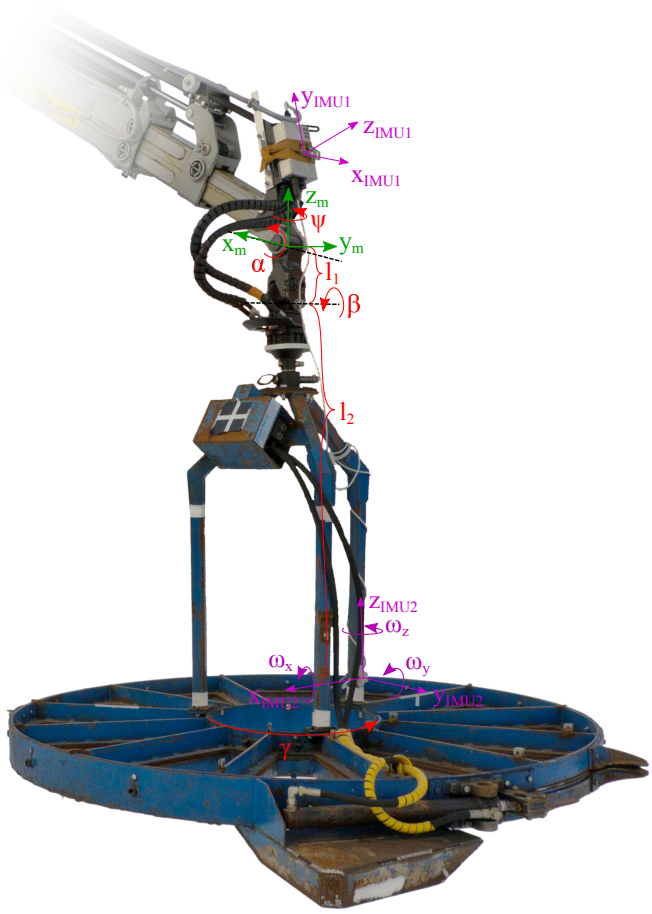


Figure 3.3. Extended Kalman filter with a two-IMU setup. Relevant dimensions, axes, and angles are included in the figure.

derived nonlinear system model is

$$
\begin{aligned}
x_1^- &= x_1 + \Delta t x_4 \\
x_2^- &= x_2 + \Delta t x_5 \\
x_3^- &= x_3 + \Delta t \big(u_6 - \sin(x_2)x_4 - \cos(x_1)\cos(x_2)u_4\big) \\
x_4^- &= x_4 + \Delta t \Big(- b_\alpha / l_3^2 x_4 + \sin(x_1)\cos(x_1)u_4^2 + \big(- \cos(x_1)u_2 \\
&\qquad - \sin(x_1)(u_3 + g) + 2l_2 \sin(x_2)x_4 x_5 + l_2 \cos(x_1)\sin(x_2)u_5 \\
&\qquad + 2l_2 \cos(x_1)\cos(x_2)x_5 u_4\big) / l_3 \Big) \\
x_5^- &= x_5 + \Delta t \Big(- b_\beta / l_2^2 x_5 + \Big(\cos(x_2)u_1 + \sin(x_1)\sin(x_2)u_2 - l_3 \sin(x_2)x_4^2 \\
&\qquad - \cos(x_1)\sin(x_2)(u_3 + g) - \sin(x_1)\big(l_1 \cos(x_2) + l_2\big)u_5 \\
&\qquad - \sin(x_2)\big(l_1 - l_3 \cos(x_1)^2\big)u_4^2 - 2l_3 \cos(x_1)\cos(x_2)x_4 u_4 \Big) / l_2 \Big),
\end{aligned}
\tag{3.27}
$$

where $x_1, \ldots, x_5$ are the five scalar components of the state vector $\mathbf{x}$ in Equation (3.25), and $u_1, \ldots, u_6$ are the six scalar components of the control input vector $\mathbf{u}$ in Equation (3.26). The minus sign as a superscript $(\cdot^-)$ indicates the predicted estimate. Note that time step indices are omitted from the equations to make them simpler. In Equation (3.27), l_1 and l_2 are lengths in the tool (see Figure 3.3), and the third length $l_3 = l_1 + l_2 \cos(x_2)$. Moreover, g is the magnitude of gravity, and b_α and b_β are the friction parameters for the angles α and β, respectively (see Publication III for the parameter values used).

The measurement model is derived using a slightly modified EKF model (compare with Equation (2.15)), where the control input $\mathbf{u}$ is also included in the measurement model h,

$$
\mathbf{y}_k = h_k(\mathbf{x}_k, \mathbf{u}_k, \boldsymbol{v}_k), \tag{3.28}
$$

where $\mathbf{x}_k$ is the state and $\boldsymbol{v}_k$ is the measurement noise at time index k similar to Equation (2.15). This modification is necessary, since the angular velocity of the boom tip in u_4 affects measurement in the dynamic model used here. The EKF measurements are the two angular velocities of the tool IMU ($\mathbf{y} = [\omega_x \;\; \omega_y]^\mathsf{T}$). They are modeled with the following nonlinear measurement model (y_1^- is a prediction for ω_x and y_2^- for ω_y):

$$
\begin{aligned}
y_1^- &= \big(x_5^- + \sin(x_1^-)u_4\big)\sin(x_3^-) + \big(\cos(x_2^-)x_4^- - \cos(x_1^-)\sin(x_2^-)u_4\big)\cos(x_3^-) \\
y_2^- &= \big(x_5^- + \sin(x_1^-)u_4\big)\cos(x_3^-) - \big(\cos(x_2^-)x_4^- - \cos(x_1^-)\sin(x_2^-)u_4\big)\sin(x_3^-).
\end{aligned}
\tag{3.29}
$$

3.6 Head Pose Estimation for AR in a Forest Machine Cabin

In Publication IV, an IMU and a machine vision camera are used together to estimate the head pose of the forest machine operator in the forest machine cabin. The head pose must be measured accurately and quickly in an augmented reality user interface where visual information is augmented in the operator's visual field using a head mounted display. The delay and mismatch that such augmentation causes between visual and vestibular cues to motion represent one of the most difficult challenges in virtual environment system designs (Badcock, Palmisano, & May, 2015, p. 73).

A DFK 41AU02 color camera (Imaging Source, 2007)[20] was integrated with an MPU-6050 IMU sensor (InvenSense, 2013) on the operator's helmet to measure the position and orientation of the helmet in real time. If the helmet would also include an augmented reality display, it could show the augmented-reality visualization correctly with respect to the cabin's pose in the forest. A software was built for a linux PC to collect time-synchronized raw-data from the camera and IMU sensors to estimate the head pose and to draw the augmented reality image in real time.

For head pose estimation, an EKF was designed which uses a state vector

$$\mathbf{x} = \begin{bmatrix} {}^n_bC_{11} & {}^n_bC_{12} & {}^n_bC_{13} & {}^n_bC_{31} & {}^n_bC_{32} & {}^n_bC_{33} \end{bmatrix}^\mathsf{T}, \tag{3.30}$$

where the last three state variables are similar to the first three state variables of the DCM IMU algorithm (see Section 3.3), but the other three state variables include the first row of the DCM matrix in Equation (3.14) instead of the bias states. Since this EKF is used in cascade with the DCM IMU, the gyroscope bias states need not be estimated again. Instead, biases estimated by DCM IMU (b_x, b_y, b_z) can be subtracted from the gyroscope measurements while generating the control input vector

$$\mathbf{u} = \begin{bmatrix} \omega_x - b_x & \omega_y - b_y & \omega_z - b_z \end{bmatrix}^\mathsf{T}. \tag{3.31}$$

In the prediction step of the proposed EKF, the following nonlinear system model (compare with Equation (3.18)) is used to predict state from the previous time step to the current one:

$$f_k(\mathbf{x}_k, \mathbf{u}_k) = \mathbf{x}_k + \begin{bmatrix} \Delta t[{}^n_b\mathbf{C}_1\times]_k \\ \Delta t[{}^n_b\mathbf{C}_3\times]_k \end{bmatrix} \mathbf{u}_k, \tag{3.32}$$

where Δt is the sampling interval and $[{}^n_b\mathbf{C}_1\times]_k$ is a skew-symmetric matrix generated from the first three state variables at time index k,

$$[{}^n_b\mathbf{C}_1\times]_k = \begin{bmatrix} 0 & -{}^n_bC_{13} & {}^n_bC_{12} \\ {}^n_bC_{13} & 0 & -{}^n_bC_{11} \\ -{}^n_bC_{12} & {}^n_bC_{11} & 0 \end{bmatrix}_k, \tag{3.33}$$

[20] The DFK 41AU02 color camera used in this thesis, manufactured by Imaging Source, is explained in detail in Section 4.2 and also used in Section 4.3.

and $[{}^n_b\mathbf{C}_3\times]_k$ is already defined in Equation (3.19).

This filter contains two measurement equations, the first is the same measurement as in DCM IMU (see Equation (3.20)), which is run at each filter iteration, since the frequency of IMU measurements is one magnitude higher than the camera image capture rate. The second measurement equation is run whenever there is a new image captured by the camera, in which the ArUco library marker detection and pose estimation (Garrido-Jurado et al., 2014; Muñoz-Salinas & Garrido-Jurado, 2016) succeed in obtaining a pose estimate. This measurement model is also linear and simple, since the rotation measurement is fed into the EKF as a DCM matrix ${}^n_b\mathbf{C}_m$ computed from the ArUco pose estimate. The measurement vector is thus

$$\mathbf{y} = \begin{bmatrix} {}^n_bC_{m,11} & {}^n_bC_{m,12} & {}^n_bC_{m,13} & {}^n_bC_{m,31} & {}^n_bC_{m,32} & {}^n_bC_{m,33} \end{bmatrix}^\mathsf{T}, \tag{3.34}$$

and a simple measurement model $\mathbf{H} = \mathbf{I}_6$ can be used to predict state to the measurement space. Note that ${}^n_b\mathbf{C}_m$ represents a rotation from the body-fixed frame to the navigation frame, so the rotation from the camera frame (where the pose estimation of ArUco occurs) to the body-fixed frame must be calibrated and taken into account in the computation.

Covariance matrices for state prediction and measurement are tuned manually to static values on the main diagonal (no correlation between states) as explained in Publication IV.

Finally, similar to the DCM IMU in Section 3.3, the results of this filter must also be normalized at the end of each iteration. This normalization could be achieved in a similar way to the DCM IMU for both DCM rows separately, but, in addition, the first and last row vectors must also remain orthogonal. Therefore, a renormalization method (Premerlani & Bizard, 2009) is used. In this method, the first and the last row vectors of DCM, ${}^n_b\mathbf{C}_1$ and ${}^n_b\mathbf{C}_3$, respectively, should be perpendicular, which means that they have a zero dot product. According to Premerlani and Bizard (2009), the value from the dot product is the amount of error which is reduced by cross-coupling using

$$\begin{aligned} {}^n_b\mathbf{C}'_1 &= {}^n_b\mathbf{C}_1 - \frac{{}^n_b\mathbf{C}_1 \bullet {}^n_b\mathbf{C}_3}{2}\,{}^n_b\mathbf{C}_3 \\ {}^n_b\mathbf{C}'_3 &= {}^n_b\mathbf{C}_3 - \frac{{}^n_b\mathbf{C}_1 \bullet {}^n_b\mathbf{C}_3}{2}\,{}^n_b\mathbf{C}_1, \end{aligned} \tag{3.35}$$

where ${}^n_b\mathbf{C}'_1$ and ${}^n_b\mathbf{C}'_3$ are the perpendicular versions of ${}^n_b\mathbf{C}_1$ and ${}^n_b\mathbf{C}_3$, respectively. Finally, the perpendicular versions must be normalized to guarantee that they remain unit vectors. Whenever the full DCM matrix is required, the second-row matrix can be computed from the last and first rows using a cross product. The effect of normalization must also be projected on the covariance matrix in a similar manner to the DCM IMU algorithm in Section 3.3. This is explained in detail in the Appendix of Publication II.

4. Exteroceptive Perception

Exteroceptive perception enables the robot to perceive the world around itself. This chapter surveys the developed exteroceptive perception methods and their sensor setups. First, the three actuated laser scanner configurations presented in Publications I, VI, and VII are explained in detail. After that, the machine vision instrumentation developed in Publication V for a point cleaning tool to detect young spruce trees is presented. In the sensor setups, the practical hardware innovations and industrial applicability are elaborated. After the sensor setups, the exteroceptive perception methods are proposed. These include real-time machine vision for young spruce detection in Publication V, young tree detection and species classification in Publication VI, and tree stem and ground model estimation in Publication VII.

4.1 Actuated 2D Laser Scanners for 3D Mapping

Actuating a laser scanner offers an affordable means of increasing the FoV of the sensor. In case of a low-cost single-beam 2D laser scanner, the actuation can provide measurements from the third dimension allowing a 3D view of the environment instead of using more expensive 3D lidars. In their comprehensive review Raj, Hanim Hashim, Baseri Huddin, Ibrahim, and Hussain (2020) report mechanical, MEMS, and solid state scanning mechanisms for building a 3D lidar, that exist in previous research and commercially available sensors. In this work, I focus on the mechanical solutions since neither MEMS nor solid-state devices were yet available during the research for this work. Mechanical solutions include mostly actuated 2D scanners and more complex scanning patters, such as the Risley scanner, which uses a pair of revolving prisms to actuate the laser beam in Lissajous curves (Vuthea & Toshiyoshi, 2018). The 2D scanners have been actuated in various different configurations either in continuous rotations or periodic oscillations. These patterns are well covered in the work by Wulf and Wagner (2003) and this work builds on top of their

results.

In this work, three different methods to actuate 2D laser scanner were tested in a forest environment. These methods are 1) a crane-boom-mounted vertical 2D scanner (used in publications I and IV), 2) a swiveling 2D scanner (used in Publication VI), and 3) rotating tilted 2D scanners (used in Publication VII). Each of these methods is able to build an accurate 3D point cloud from the working environment. They all offer advantages and disadvantages, and they are best suited to certain tasks.

Crane Boom Mounted Vertical 2D Scanner

One of the simplest and most straightforward methods to actuate 2D scanner to measure a 3D point cloud is to yaw it perpendicular to the scanning plane (Wulf & Wagner, 2003). In Publication I, the scanner is mounted vertically on the side of the forestry crane to observe the two targets also attached to the side of the crane. As noted earlier, most of the light pulses are then free to travel further into the environment, providing range measurements from it on the vertical plane aligned towards the crane boom. As the crane is repeatedly moved from side to side when it is normally used in many different tasks, the environment can be measured simultaneously. Here, the evident benefit is that the same sensor can be used to measure both the posture of the flexible forestry crane (see Section 3.1) and also the environment, thus offering an extremely low-cost solution for environmental data collection.

The largest disadvantage of this instrumentation is the limitation of the data to one vertical line in the direction of the crane. Thus, only data from this single cut plane of the environment are obtained at a time. However, this could easily be avoided by using a modern, low-cost multi-beam lidar, for instance, a 16 beam Velodyne Puck Lite (Velodyne Lidar, 2019b), a 32 beam Ouster OS1 (Ouster, 2021), or a 40 beam Hesai Pandar40P (Hesai, 2020), instead of the currently used SICK LMS221 2D scanner (SICK AG, 2006).

The other significant drawback is that a large share of the data is lost towards the ground and the sky, since, in the yawing scan, the point cloud is the densest near the z axis pointing upwards (Wulf & Wagner, 2003). On the other hand, in this yawing scan configuration, the point cloud is uniform, and its density is almost equal near horizontal plane if the yawing speed is constant. This is beneficial in this forest machine case, since targets such as trees are commonly nearer the horizontal plane than the z axis. Furthermore, an excessively dense point cloud at the top and bottom can be decimated by discarding unnecessary points.

An example of a 3D point cloud collected in Publication I is shown in Figure 4.1. The example point cloud has been decimated to provide equal density of points per solid angle. This configuration is best for measuring

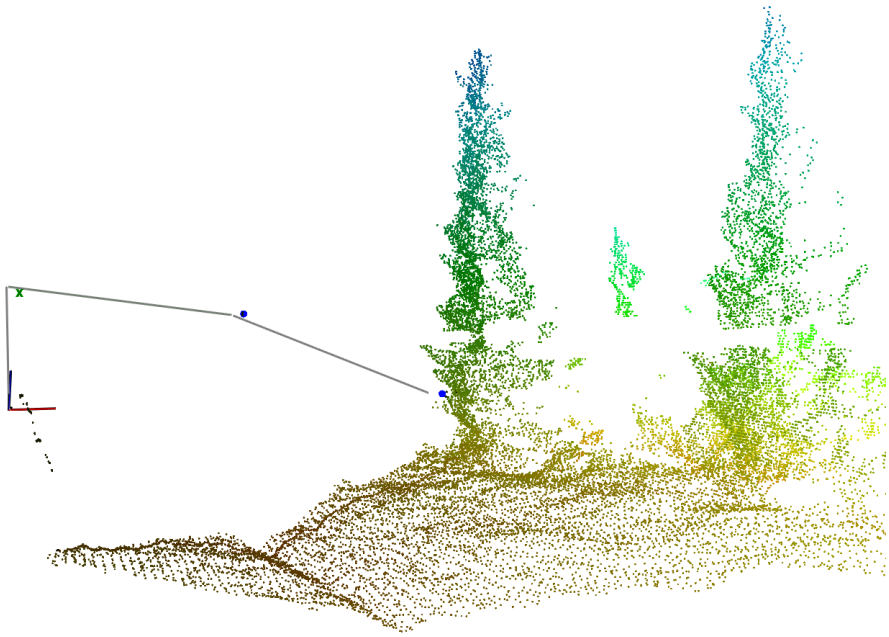


Figure 4.1. A point cloud accumulated over 125 seconds when the crane was rotated from right to left towards a spruce tree. The crane is drawn as a gray line, the laser scanner is located at the green cross, and the targets are the two blue dots. Range measurements associated with Targets 1 and 2 are removed from the point cloud, and the cloud is colorized in HSL color space such that hue indicates the height of a point, lightness the range of a point from the scanner, and saturation is set to one.

crane posture, but it can also provide some valuable information from the surrounding forest at the same time. The 3D model updates slowly, so the method is not viable as the only 3D sensor for a forest robot. However, the crane tip position is directly drawn in the same coordinates as the 3D point cloud, so it would be straightforward to combine the crane position with the point cloud. This would be beneficial for combining existing forest machine data with point cloud data. Therefore, this instrumentation could be usable for collecting individual tree-level information from the forest plot under operation.

Swiveling 2D Scanner

To improve point cloud density from the previous example involving a simple rotation around one principal axis, some more complex rotations have been considered. For example, Nagatani, Tokunaga, Okada, and Yoshida (2008) propose a 30-degree-downward-tilted horizontal 2D scanner which is yawed horizontally. They claim two advantages for this 30° tilted configuration. The first is that it provides better distribution of the scan points and avoids the dense areas near the rotation axis, as found in the previous section. The second benefit is that it increases the scanning speed

or scanning density, since rotation can be performed faster to achieve a similar density with a tilted scanner.

To enhance the possibility of adjusting the inclination of the 2D laser scanner, some authors have proposed that the scanner should be mounted on a servo-controlled gimbal unit or a pan-tilt unit allowing rotation on two or more axes (e.g., Ocando, Certad, Alvarado, & Terrones, 2017; Khurana & Nagla, 2020). These setups allow more complex continuous Lissajous-like scanning patterns (Anderson & Clayton, 2014), which enable high resolution scanning in less time than other scanning types.

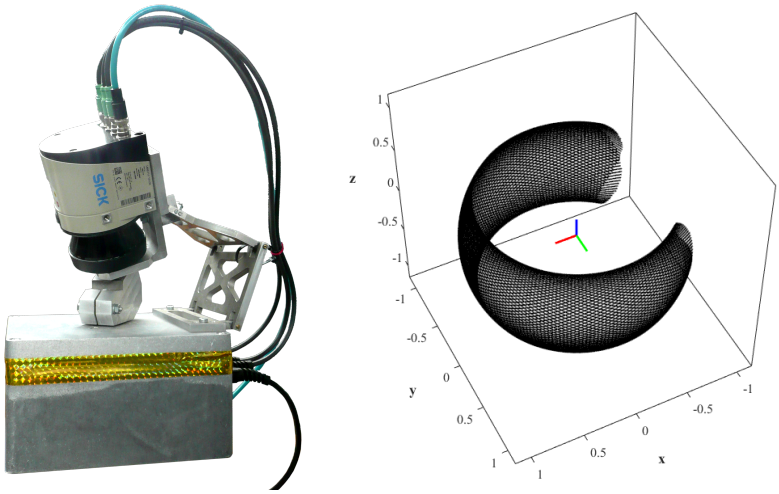
Usually, these complex motions require the measurement of multiple motors and multiple joint angles. However, Yoshida, Irie, Koyanagi, and Tomono (2011) have shown that a swiveling type gimbal motion for a horizontal 2D laser scanner can be achieved with only one motor. They also propose a secondary motor to modify the density of the point cloud on the regions of interest.

In our laboratory[21], the design by Yoshida et al. (2011) has been simplified, and a swiveling mechanism with only one motor has been built featuring a 30° wedge rotated by a motor built inside a metal box underneath (see Figure 4.2a). A SICK LMS111 2D scanner (SICK AG, 2009) has been mounted upside down on a gimbal structure with bearings on the bottom and back, which allows it to rotate on the gimbal when the wedge is rotated. The angular position of the motor is measured with an angular position sensor on the motor axis inside the bottom box in Figure 4.2a. The arm at the back of the scanner keeps the heading of the scanner fixed. The resulting motion is a similar swiveling motion to that proposed by Yoshida et al. (2011) and later named swiveling motion by Oberländer, Pfotzer, Roennau, and Dillmann (2015).

There are some benefits to the swiveling motion. Firstly, a large FoV of nearly $60° \times 270°$ is achieved, with the SICK scanner measuring 270° FoV on a plane. Secondly, the full FoV can be scanned in a short period of time, providing quite uniform point density. For example, a rotation velocity of 30 revolutions per minute (RPM) was used in Figure 4.2b. Thirdly, there is no need to feed the scanner cabling through a continuously rotating joint (using, e.g., a slip ring), as required in other continuously rotating setups, since the heading of the scanner is fixed.

One of the disadvantages of the current swiveling 2D scanner prototype is that the vertical FoV is limited to ±30°, which is slightly too little in the forest case, as can be observed in the example projections of range and intensity data in figures 4.3a and 4.3b, where some of the tree tops and bottoms are cropped out of the image. The sensor also possesses limited FoV in the horizontal direction, as it can not see anything to the rear. Therefore, the sensor is not viable as a single sensor for a robot moving

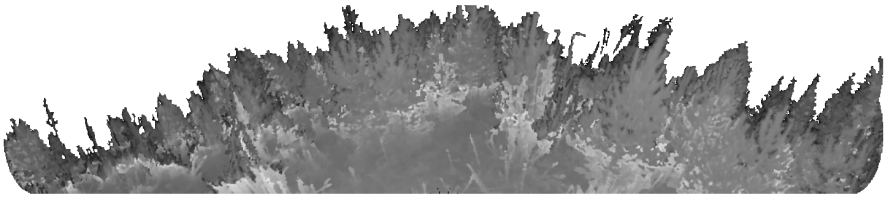
[21]The sensor was designed and built by Matti Öhman and Tapio Leppänen.



(a) Swiveling scanner prototype **(b)** Simulated point cloud

Figure 4.2. Swiveling 2D scanner and a simulated point cloud (range set to 1) to show the distribution of points when the swiveling mechanism is rotated at 30 RPM. The scanner is placed at origin, and axis directions are drawn on the figure with red (x, forward), green (y) and blue (z) lines.

(a) Range data (brighter is further away) projected on a sphere



(b) Reflection intensity data projected on a sphere

Figure 4.3. The swiveling scanner achieves a 270° horizontal and 60° vertical FoV. The data are projected on spherical coordinates around the forward pointing axis such that the x coordinate represents longitude from -135° to 135° and the y coordinate represents latitude from -30° to 30° with respect to the center of the images. The data were collected for Publication VI in a young mixed species forest.

around in a forest.

Moreover, a challenging drawback of the setup is the swiveling motion itself, which produces accelerations that cause low frequency vibrations in the system. The swiveling motion produces periodic accelerating and decelerating sideways motions as the gimbal oscillates. These vibrations

are difficult to avoid in such swiveling types of motion, since the scanner always has at least some mass. To observe the trees from above in Publication VI, the swiveling scanner was mounted on top of a pole, which amplified the vibrations, causing considerable motion blur and inaccuracies in the 3D point cloud.

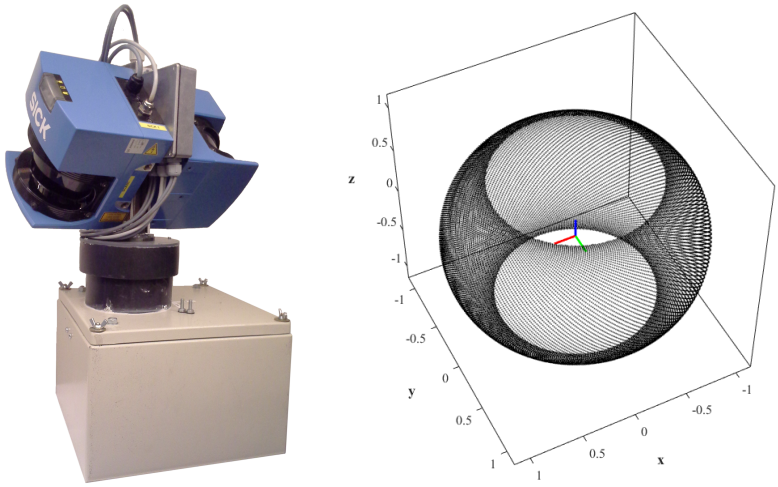
The swiveling scanner is best for quickly scanning a limited area of 3D space to the front and side of the robot (see Figure 4.2b). The data can be easily combined with a color camera, as shown in Publication VI, to perform sensor fusion between the lidar and the camera. This sensor configuration is beneficial when the limited FoV is sufficiently large for the task. However, as the sensor wobbles as the gimbal oscillates, the sensor should be mounted tightly next to the main body of the robot. The use case in Publication VI, where the sensor is located on top of a high pole, is a good example of how a wobbling sensor should not be mounted.

Rotating Tilted 2D Scanners

Through an iterative process of trial and error in the research for this thesis, it became clear that a vertical FoV of at least 90° is required to see tree tops and the ground surface underneath the trees in the forest environment. To build a configuration from an actuated 2D scanner which simultaneously optimizes the uniformity of the resulting point cloud, allows usage of all measurements, provides a sufficiently large FoV, and avoids the noted drawbacks of the swiveling motion, a rotating 45° sideways-tilted scanning configuration was found to be the best alternative. The idea is similar to that already proposed by Nagatani et al. (2008), but the tilting is sideways instead of downwards nodding.

With a 45° sideways tilted scanning plane, a $90° \times 360°$ FoV can be achieved with highly uniform point cloud density allowing the use of all range measurements (see Figure 4.4b). If another similar scanner is also mounted such that the scanners face backwards on the same pole (see Figure 4.4a), the center of mass of the rotating part can be tuned to be the exact center of the rotating axis. This removes most of the forces causing vibrations found with the swiveling scanner. Naturally, this also doubles the data rate.

Two SICK LMS200 2D scanners (SICK AG, 2006) were used in this configuration. They can measure 75 scans per second with 180° FoV on a plane at a 1° angular resolution. The first prototype of the rotation mechanism for this structure was used in Publication VII. It utilized a wireless data transfer device by Moxa to transfer data from scanners mounted on the rotating part. However, random and unknown delays were introduced in the wireless data link. In addition, some packet loss was encountered. To correct these issues, a second prototype, shown in Figure 4.4a, was built with the same scanner configuration. It features



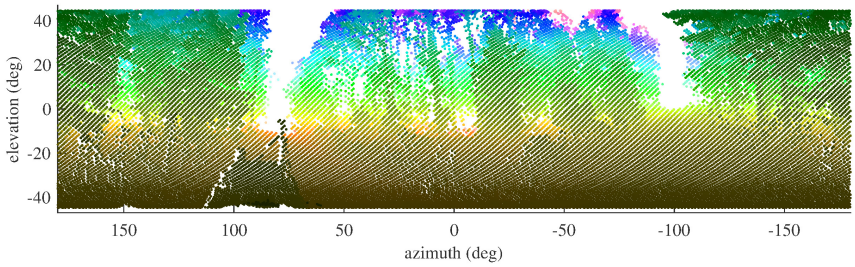
(a) Rotating scanner prototype **(b)** Simulated point cloud

Figure 4.4. The rotating tilted 2D scanner and a simulated point cloud (range set to 1) to show the distribution of points when the scanner is rotated at 30 RPM. The rotating shaft between the scanners is placed at origin, and axis directions are drawn on the figure with red (x), green (y), and blue (z) lines. Only the points of the first scanner are drawn to keep the image clearer. Scanners can be timed such that the other scanner aims between the lines scanned by the first sensor, doubling the effective resolution.

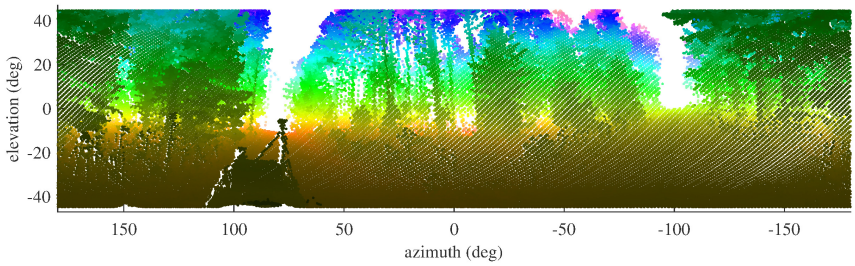
a slipring for data and power connections between the bottom and the rotating parts. The new version also contains a more accurate encoder (ASM Posimag incremental encoder and a magnetic ring with a reset pulse once per revolution) for measuring the angular position with a resolution of 0.02°.

One laser scanner provides 150 scans distributed evenly on the full horizontal 360 degree FoV in 2 seconds when the motor is rotated at 30 RPM. Using both of the laser scanners mounted back to back (as in Figure 4.4a), the data rate can be doubled and the full FoV measured in a second. Example scans using a) only one, or b) both lidars are shown in Figure 4.5. Figure 4.5c on the other hand accumulates 5 consecutive rotations during a 10 second interval to show that longer time period can be used to accumulate a higher resolution point cloud assuming that the sensor position and orientation are known.

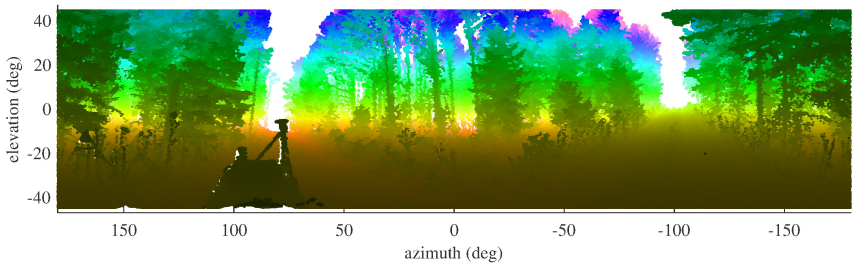
This rotating scanner can measure in all relevant directions for a forest machine, and the resulting evenly distributed 3D point cloud is updated once every second (assuming that both scanners are measuring and they are rotated at 30 RPM). A similar point distribution is also available in RobotEye RE08 3D lidar by Ocular Robotics, but only in 70° vertical FoV (Wood & Bishop, 2012). In addition to providing a large vertical FoV, the selected 45° tilt angle is beneficial for later data processing. For example, in Publication VII, the horizontal ground surface and vertical tree stems



(a) One scanner, 360° rotation in 2 seconds at 30 RPM



(b) Two scanners, 360° rotation in 2 seconds at 30 RPM



(c) Two scanners, 5 full rounds in 10 seconds at 30 RPM

Figure 4.5. Examples of rotating 3D laser scanner data in a forest when the scanner is mounted on an ATV as shown in Figure 1.1 earlier. The data is visualized as a panorama image centered on the middle of the scanners and colored in HSL color space similarly as the point cloud in Figure 4.1.

can be detected from individual 2D scan lines. This can be achieved by assuming that the rotating axis primarily points upwards, such that the 45° tilted scan lines mostly see both edges of the tree stems and end up with ground-surface-associated measurements at the bottom. This rotating tilted-2D-scanner configuration is feasible as a forest robot's main sensor for tree detection, mapping, navigation, and surround monitoring purposes. However, it is not viable for measuring a focused high resolution point cloud on a narrow FoV.

4.2 Cleaning Tool Mounted Camera for Young Spruce Detection

In Publication V, an automated point-cleaning robot with a machine vision camera is used to detect young spruces. To provide the remote operator and the robot with an unobstructed view under the cleaning tool, the camera must be placed near the bottom of the central hole of the tool (see Figure 4.6a). However, the camera should not be directly mounted onto the bottom of the cleaning tool, since the tool is operated by lowering it through the vegetation to remove foliage around the target tree. If the camera were not moved away, it would become dirty, optics would be scratched, and it could even break if it were in direct contact with the ground, trees, and other obstacles. To solve the problem, a camera-lift system is proposed (see figures 4.6 and 4.7) which removes the camera from danger by enclosing it in a metal box during the point cleaning operation.

In addition to computer control, a human operator can also use the camera to aim the cleaning tool at the target tree whose surroundings are being cleaned. The operator might forget to close the camera before lowering the cleaning tool on top of a tree, thus causing the camera to collide with the tree. Therefore, the camera lift also requires a safety mechanism that quickly hides the camera before it collides with an object. For this task, an ultrasonic range finder (A in Figure 4.7a) has been added to the same box next to the camera lens (B in Figure 4.7a) to automatically close the box when any object is detected too near the camera lens.

The low-cost ultrasonic range finder used in the study is a Devantech SRF08 (Coe, 2001). It has a wide beam width of about 55°, and it measures up to 16 sequential ranges from 3 cm to 6 m during the recommended 65 ms measuring time (Koval, Vaňuš, & Bilík, 2016). In this case, only

(a) The cleaning tool with the camera lift **(b)** The camera is enclosed and safe

Figure 4.6. The cleaning tool features a hole at its center to allow it to clean around the target tree. The camera-lift mechanism can be closed when the cleaning operation is performed to allow the target tree to stand freely in the center.

(a) The camera lift system **(b)** An example view of the camera

Figure 4.7. An ultrasonic sensor (A) and a machine vision camera (B) are mounted inside a camera lift mechanism, which retracts the sensors inside the box using a motor-actuated lever (C) when the ultrasonic sensor detects an obstacle in front of the camera. The camera lift can be opened and closed by a human operator or the automated system controlling the crane. The camera has a wide FoV under the cleaning tool to observe nearby young spruce trees.

the shortest range to any object is considered, and the measuring time is shortened to 25 ms (equivalent to about 4 m maximum range) to reduce the dead time. The decision limit to react to any object has been set to 35 cm. To avoid false actions, two sequential detections of short ultrasonic range measurements are required. Thus, the system has a dead time of less than 50 ms before it reacts to any object nearer than 35 cm. After detection, the motor-actuated lever pulls the camera back and encloses it in the box within a second. This should be fast enough because the tool is lowered slower than it takes for the camera lift to close.

The camera (B in Figure 4.7a) is a DFK 41AU02 low-cost color camera with a sensitive ½ inch Sony CCD ICX205AK sensor (Imaging Source, 2007). Data are collected in a raw format at 15 frames per second (FPS) using a resolution of 1280×960 pixels. The raw data are read by the boom tip computer and broadcast in real time over a wired Ethernet network installed on the crane. It is necessary to observe simultaneously a wide area below the tool to detect nearby spruce trees effectively, so a lens with a focal length of 6.5 mm was used, providing approximately 90° FoV for the shorter image axis. An example image is presented in Figure 4.7b.

Since the camera lift closes the camera inside the metal box, the amount of light entering the sensor changes rapidly from bright sunlight to pitch black. This renders most of the built-in integrated automatic exposure and gain controllers normally present in machine vision cameras unusable. The gain controller in the camera used in the study was far too slow to adapt to bright sunlight after the camera had been kept in the box for a prolonged period. However, fixed values for the parameters could not be used either, since the amount of sunlight can change drastically during

normal operation (e.g., a shift from direct sunlight to shade).

To solve the problem, the built-in automatic gain and exposure controller of the machine vision camera was first disabled. Second, a custom external gain and exposure controller was implemented to the software reading the camera data. Instead of implementing accurate but computationally expensive state-of-the-art gain and exposure controller (see, e.g., L. Zhang et al., 2020; Bégin & Hunter, 2022), a simple and computationally efficient solution was searched. The implemented controller is a simplified and modified version of the method used by Fowler (2005) and many others cited in his work. The implemented controller reads the intensity values from the captured image data and sets the fixed camera's exposure and gain values through its control interface for each frame. The controller is designed as a simple P-controller which stabilizes the count of maximum intensity valued pixels $\sum_{i \in N_{pixels}} (I_i \geq 254)$ against a reference count of overexposed pixels, $N_{overexposed}$, thus decreasing the exposure and gain in the presence of too many overexposed pixels, and vice versa.

In the developed controller, exposure and gain are controlled together such that the exposure time remained under a fixed upper bound threshold value ($< 1/15$ s) to maintain the frame rate at a constant maximum of 15 Hz. When it is necessary for the controller to reduce the brightness of the image, gain is first reduced to zero. After that, physical exposure time is shortened to reduce brightness. On the other hand, when the controller increases the brightness, the exposure time is first increased to the upper bound limit, and then gain is increased if more brightness is required.

The controller also uses the information about the camera lift position, and thus the exposure and gain control can be enabled only when the camera observes the environment outside the box. The gain and exposure values can then be locked when the camera lift is moving or when the camera is enclosed inside the box. The image data is later used (as explained in Section 4.3) to detect young spruce trees under the point cleaning tool and thus aim the crane over the trees automatically.

This kind of controller to limit the amount of overexposed pixels while using minimally gain and maximally exposure time can be practically implemented to any machine vision solution in which the camera allows manual setting of exposure and gain parameters before capturing each frame. Counting the amount of overexposed pixels is fast and can be done from raw images directly without decoding (e.g., demosaicing) data. The implemented method was not optimal, but it was sufficient for the task at hand.

4.3 Real-Time Machine Vision for Young Spruce Detection

In Publication V, a color camera[22] is used to search for young spruce trees among other vegetation in real time. The real-time operation was required for visual servoing the forestry crane in the point cleaning task. The camera lift system explained in Section 4.2 is utilized to move the camera out of the way and to enclose it in a box when the point cleaning tool is cleaning the surroundings of a target spruce tree.

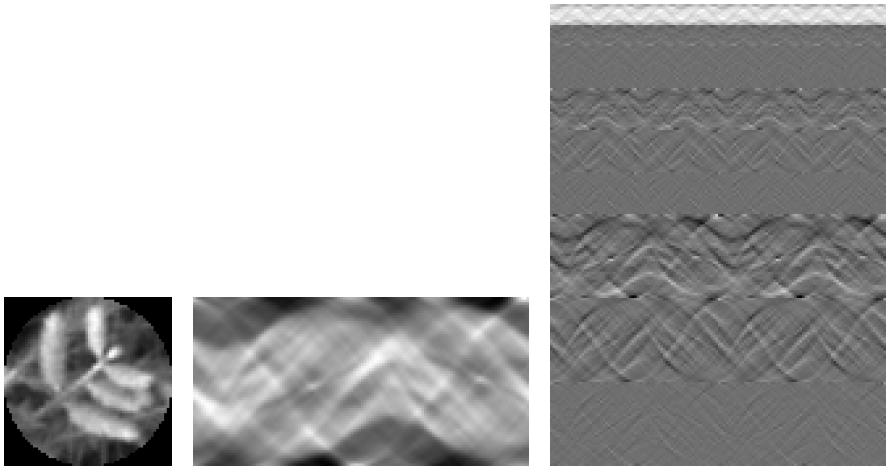
The spruce search utilizes rotation-invariant texture features which are combined with color features. These rotation-invariant texture features are computed from an intensity image. In the process, the image is first divided into smaller blocks that are processed and classified individually. Block sizes of 32x32, 64x64, and 128x128 pixels were compared in Publication V. The blocks always overlap each other by half the block size; i.e., when using blocks with 32x32 pixels, they overlap their neighboring blocks by 16 pixels each. This results in a texture analysis for a grid with nodes 16 pixels apart.

The analysis of each block is based on a rotation invariant texture analysis method by Jafari-Khouzani and Soltanian-Zadeh (2005) that uses Radon and Wavelet transforms. Rotation invariant features are used in this task, since the viewpoint is from above and young spruce trees are roughly radially symmetric. A circular windowing function is used to shape the image block to be radially symmetric. An example block of 64x64 pixels is shown in Figure 4.8a. The selected block is transformed using a standard Radon transform (see Figure 4.8b), which is further wavelet transformed (see Figure 4.8c).

Jafari-Khouzani and Soltanian-Zadeh (2005) proposed that instead of using a normal 2D-wavelet transform, a translation invariant wavelet transform (J. Liang & Parks, 1996) should be used in the x-axis and normal wavelet transform in the y-axis. For wavelet transformation a simple Haar wavelet was used with three wavelet levels causing nine sub-bands and a residual. In the results (see Figure 4.8c), the high frequency bands are at the bottom, the lower frequency bands above them, and the residual at the top. Finally, the texture features are calculated using the mean of square roots of absolute values for each sub-band in the wavelet transform. These features were used because Jafari-Khouzani and Soltanian-Zadeh (2005) showed that they provide better results compared to traditional energy and uniformity measures.

In addition to texture, color features were also combined to aid the detection of spruce among other vegetation. As found earlier in Section 2.4, in order to provide a suitable color space for use with natural images, the fixed linear transformation defined in Equation (2.24) was employed to

[22]The DFK 41AU02 color camera used in this thesis, manufactured by Imaging Source, is explained in detail in Section 4.2 and also used in Section 3.6.

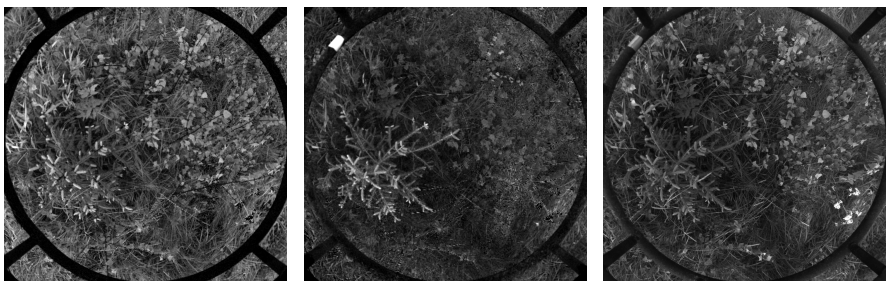


(a) Image block **(b)** Radon transformation **(c)** Wavelet transformation

Figure 4.8. The proposed texture feature computation procedure has three steps: a) an intensity image block of size 64x64 pixels with a windowing function, b) a Radon transform computed from (a), and c) a wavelet transform computed from (b)

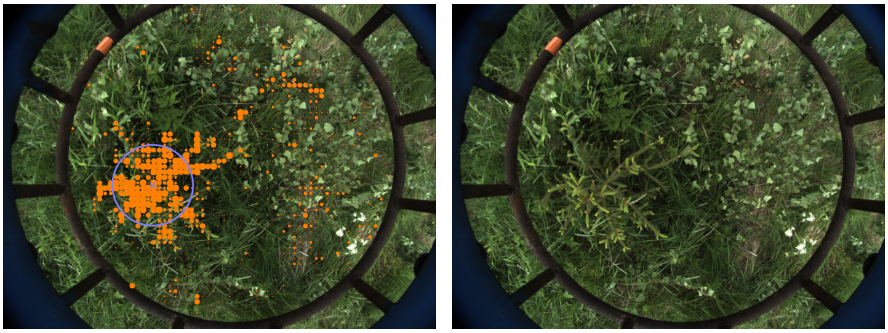
transform an RGB image to excessive green (EG), redness-blueness (RB), and intensity (I) channels (see Figure 4.9). Furthermore, as it separates image intensity (I channel), it is easy to reduce the effect of shadows by normalizing intensity in other channels by dividing each EG and RB channel pixel-wise by the intensity channel I. These new color features $EG' = EG/I$ and $RB' = RB/I$ were averaged around the same image blocks used in the texture analysis.

A k nearest neighborhood (k-NN) classifier with Euclidean distance was implemented to classify each image block into a spruce or non-spruce class based on a set of texture and color features associated with the block. The selected classifier is simple to implement and often regarded as a sort of a baseline classifier as found in Section 2.4. Since nearly any modern classifier could be utilized for this task, the simplest solution was selected



(a) Excessive green (EG) **(b)** Redness-blueness (RB) **(c)** Intensity (I)

Figure 4.9. Example color channels of a frame in EG-RB-I color space defined in Equation (2.24)



(a) Detection result **(b)** Input image for comparison

Figure 4.10. Spruce detection result with the identified tree location and fitting neighborhood drawn over the input image (b)

to demonstrate the real-time spruce detection in a forest. In the original k-NN algorithm, an unknown sample is assigned to the class most commonly represented in the collection of its neighborhood in the training data (see, e.g., Lampinen et al., 1998, p. 31). Instead of this sharp classification result, the neighborhood of size $k = 7$ was used to calculate only the number of votes $V \in \{0, 1, ..., 7\}$ for the two classes (spruce or non-spruce).

Then in the developed method, the most spruce-like area is sought from the figure using the spruce vote information around the image. The initial location is guessed by taking a median separately in the x and y directions of the image block coordinates in which there are more than a set threshold ($V_{spruce} > 3$) of spruce votes.

In the method, the actual spruce location is refined using a spruce-vote weighted average of x and y coordinates in the local neighborhood around the initial location. The circle size defining the local neighborhood is tuned to correspond to an average target tree in the image. Furthermore, the quality of the tree detection is estimated by counting the sum of spruce votes inside a circle centered to the estimated location and dividing it by all spruce votes in the image. A predefined quality threshold ($q_{spruce} > 0.2$) is required to accept the spruce detection.

The detection result was drawn on the image using orange dots (see Figure 4.10). The radius of dots corresponds to the number of spruce votes on each image block. In Publication V, the feature extraction and k-NN search were implemented using CUDA GPU computation to enable real-time capable operation.

4.4 Young Tree Detection and Species Classification

In Publication VI, the lidar and camera were used together to detect and classify young spruce and birch trees in the forest environment. The 3D li-

dar (the swiveling 2D scanner proposed in Section 4.1) and a color machine vision camera[23] were first calibrated to operate as a single sensor providing a depth estimate for each pixel on the common FoV of both sensors. To enable the use of existing camera calibration tools (e.g., Bouguet, 2004), the lidar point cloud was rendered as pairs of range and intensity images using the kinematic structure and angular position of the platform (see the swiveling 2D scanner in Section 4.1).

In the developed method, to detect young trees and classify their species, the range images are first segmented to detect tree regions. Since young trees grow close to each other, it is really challenging to segment trees from images. No suitable methods using image data only were found during the research. Not many methods existed for segmenting young trees from point cloud data either. However, a method by Jokelainen (2010), which started tree segmentation from detected tree tops, was available and was selected for the task. Segmentation is based on a flood-fill-type algorithm that operates in the following way. Starting from a seed point (assumed to belong to a tree), the method segments neighboring pixels into the same tree segment if the range difference between the current and the neighboring pixel is within the defined threshold (see Publication VI for the values used). The seed points can be selected in each tree using the highest local points on the 3D point cloud as an indicator of a possible tree (see, e.g., Jokelainen, 2010; Vihlman, 2012).

Removal of segments from the ground and smaller vegetation around the trees is performed using the assumption that a tree grows upwards. The Cartesian coordinates of the segment (computed from the depth image) are fed into principal component analysis (PCA) to find three coordinate axes in the order of diminishing variance. The axis nearest the z axis (upwards) is decided to estimate the direction of growth (i.e., the height axis), the other two form the horizontal plane. To reduce the effect of ground points, PCA axes are formed iteratively by removing the lowest points of the segment at each iteration. Finally, the distribution of the remaining higher points in the horizontal PCA plane defines a threshold window for removing surrounding points from the original segment, thus reducing the effect of ground points being associated with tree segments.

PCA is also useful for detecting segments containing only ground points (instead of a tree surrounded by ground). A segment is considered a ground-only segment,

a) if it spans a longer distance along both horizontal axes compared to the distance along the height axis, or

b) if its span is at least 60% longer along the longer horizontal axis than

[23] The NET GmbH Foculus FO442C is a color camera with a 2/3" progressive scan CCD image sensor that offers a resolution of 1392×1040 pixels and measures 12 bit raw images at 20 Hz frequency (Aegis Electronic Group, 2006).

it is along the height axis.

These rules were formed experimentally within the training samples. Since the tree segmenting had good enough success in the young forest under investigation, further improvement of segmentation was left for the future work. The developed method requires tops of the trees to be visible in the data and to be local maxima of the point cloud, so the method should be applicable in a different young forest as well as long as the trees are the highest objects on the scene.

After segmentation, tree regions were then projected onto the camera image using the transformation obtained using external stereo camera calibration between the lidar intensity image and the machine vision camera. The tree segments on the color camera images were then divided into rectangular image blocks (varying size of about 40×40 pixels). To survey which of the commonly used image features are most suitable for a young forest, various texture features from the following eight groups were computed for each image block[24]:

a) 20 descriptors of the Gray Level Co-occurrence Matrix (GLCM) (Haralick, Shanmugam, & Dinstein, 1973): contrast, correlation, energy, homogeneity and entropy in four directions.

b) 44 descriptors of the Gray Level Run Length matrix (GLRL) (X. Tang, 1998): 11 features in four directions.

c) 2 descriptors of edge frequency: the number of edge pixels per unit area using the Roberts operator (first used by Roberts, 1963) and the zero crossing Laplacian of Gaussian, also called the Marr–Hildreth algorithm (Marr & Hildreth, 1980).

d) 4 fractal dimension descriptors (B. Chaudhuri, Sarkar, & Kundu, 1993).

e) 16 statistical geometrical features (Y. Q. Chen, Nixon, & Thomas, 1995).

f) 2 descriptors of Local Binary Patterns (LBP) (Ojala, Pietikainen, & Maenpaa, 2002): mean and standard deviation of the rotation invariant LBP histogram.

g) 7 descriptors of the three-level wavelet decomposition using Daubechies db1 wavelets: three measures of energy in the wavelet approximation and detail images, and the four rotation invariant features (Porter & Canagarajah, 1997b).

h) 4 rotation invariant Gabor filter descriptors (Porter & Canagarajah, 1997b).

These features were used to classify the tree species using a two-class

[24]For GLCM and GLRL, the image was first scaled to have eight numerical levels.

Naive Bayes classifier (see the Camera subsection in Section 2.4 for details). The classification procedure contains two steps. In the first step, each image block belonging to a tree is classified individually. In the second step, each image block is given a value ±1 depending on which of the two classes has a higher density. The whole segment is then classified based on the density-weighted sum of block classes.

Since real-time computing of all the surveyed features would be infeasible in an autonomous forest machine, it was important to find a minimal set of suitable features providing the best classification results. In the extensive comparison of the features in Publication VI, a selection of eight features from the co-occurence matrix, local binary patterns, statistical geometrical features, and Gabor filter were found to produce the best classification result. In Publication VI, the classification results were similar for spruce and birch. The overall correct detection rate was 79% and the overall correct classification rate was 74%. The study demonstrates that it is plausible to combine lidar and camera measurements to detect young trees and classify their species.

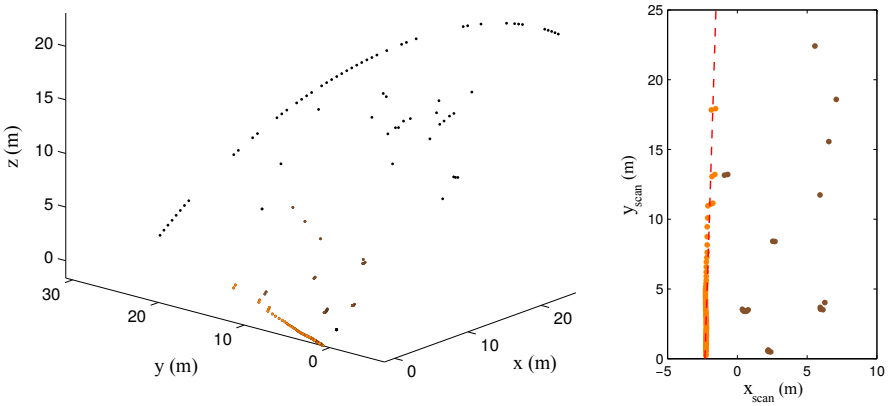
4.5 Tree Stem and Ground Model Estimation

Two rotating tilted 2D scanners[25] in the configuration proposed in Section 4.1 were used to measure and model the main stems (i.e., trunks) of trees in a forest Publication VII. To measure the main stems and model individual trees with models related to the local ground surface, both the ground surface and tree measurements must be located from the 3D point cloud. For this task, a method utilizing the 45° tilted 2D laser scanners was developed. The method assumes that tree trunks primarily grow vertically and the ground surface usually lies horizontally underneath the trees, and thus the laser profile strikes both the ground surface and tree stems at approximately a 45° angle.

Because of the 45° tilted scanning plane, ground and tree points can be segmented from individual scan lines directly without the need of combining multiple scan lines of the rotating 2D lidar (see Figure 4.11). In this scanning configuration, the ground is usually seen as a linearly arranged set of points in the left side of the laser scan , at least when the ground surface or the measuring vehicle are far less than 45° inclined. In addition to the rotating laser scanners, the setup in Publication VII also included an IMU sensor, to allow the vehicle attitude to be taken into account if the platform is over-tilted (e.g., using the method proposed in Section 3.3).

Ground surface segmentation is performed for a single laser scan by iteratively fitting a line to the 3D points at the lower part of the data (on

[25] The system integrates two SICK LMS200 2D scanners (SICK AG, 2006).



(a) A sample 2D laser scan in a scanner-centered 3D view **(b)** Scanning plane view

Figure 4.11. A sample scan from the rotating lidar is shown in 3D (a) and in the scanning plane (b). The ground points are visualized with orange dots, and the ground line is shown with a red dashed line. Brown dots visualize the common field of view between (a) and (b), and the other measurements are shown as black dots in (a).

the left side in Figure 4.11b). Laser range measurements that are near the fitted line or to the left of the line are associated as ground returns. The position of the fitted line is initially selected to be parallel to the y axis in the scanning plane and to go through the leftmost returns in the selected search window shown in Figure 4.11b.

To build the ground model, ground-associated measurements during a full revolution of the rotating lidar (see rotating tilted 2D scanners in Section 4.1) were collected in a single scanner-centered point cloud. In the ground modeling process, the point cloud is first randomly thinned to equalize the density of points in it. Then, the point cloud is split into 1×1 m x-y grid cells, after which the median height (z) is computed for each grid cell to represent the ground surface around that cell. A 1 m grid was chosen because, in practice, it proved sufficiently accurate for use in the remaining tree modeling phase to estimate the ground surface level under each tree. Furthermore, the selected resolution allowed sufficient points to be accumulated in each grid cell for the median to be robust against possible random errors in the segmentation of laser range measurements. For empty grid cells in the middle of the ground model, the height was interpolated using neighboring cells, from which the median could be computed.

Tree stem segmentation is also based on the separate segmentation of each laser scan. Publication VII proposed a method for detecting tree stem edges and tree stem returns in between the edges. Because of the lidar tilt angle, primarily vertically growing tree stems were measured at approximately 45° in relation to the tree growing direction. Thus, it can be assumed that, in the laser scan, the points are measured and organized

in an angular order from left to right. Therefore, the left edge logically appears first, after which the stem points and then the right edge of the same tree are encountered in the laser-scan vector. Similar detectors are used in many 2D mapping cases to detect trees from a laser scan (see, e.g., Bailey & Nebot, 2001; Jutila et al., 2007).

Relying solely on a single scan segmentation to the left and right edges and stem points would produce a high number of outliers caused by range measurements from obstacles such as rocks, thicker branches, or undergrowth. To improve the segmentation, a second step is performed in the tree stem search. First, a full revolution of the rotating lidar data initially classified as tree points (edges or stem points) is collected in a scanner-centered point cloud. Then tree trunks are searched for from the point cloud by computing a x-y histogram using a bin width of 0.2 m in both x and y directions. The trunks are assumed to be nearly vertical, so they should be seen as local maximums in the 2D histogram.

The number of points inside one histogram bin is highly biased by the measuring range. To separate non-tree objects from real trees, a range-dependent threshold value was used to classify histogram bins into trees or non-trees. This threshold value is modeled as a function of range with the help of a geometrical model of a solid angle (Ω) of a tree in the FoV of the rotating lidar:

$$\begin{aligned}
\Omega(r) &= 2\arctan\left(\frac{d}{2r}\left(\sin\alpha(r)+\sin\beta(r)\right)\right),\\
\alpha(r) &= \min\left(\frac{\pi}{4},\arctan\left(\frac{z_{\max}-z_0}{r}\right)\right),\\
\beta(r) &= \min\left(\frac{\pi}{4},\arctan\left(\frac{z_0}{r}\right)\right),
\end{aligned} \tag{4.1}$$

where d is the diameter and r is the range of the model tree trunk, $z_{\max}$ is the maximum height of trees where the tree trunk is visible in the forest (10 meters used in Publication VII), and z_0 is the scanner height from the ground surface.

Since the rotating lidar (see rotating tilted 2D scanners in Section 4.1) produces rather equal point density over its FoV, the detection threshold can be assumed to be proportional to the solid angle of a model tree in Equation (4.1). The constant parameter defining the scaling between the solid angle model and the threshold is tuned based on the data. The histogram of points and the threshold are shown in Figure 4.12.

The approximate tree locations are then searched for from the x-y histogram around the accepted bins. To avoid finding the same tree more than once (e.g., if the same trunk is split into two neighboring cells), histogram bins closer than 1 m to each other are assumed to belong to the same tree. Thus, only the local maximum around every candidate (red dots) which exceeded the modeled threshold function (black curve) in Figure 4.12 is determined as a tree (green circles). This allows only the largest histogram

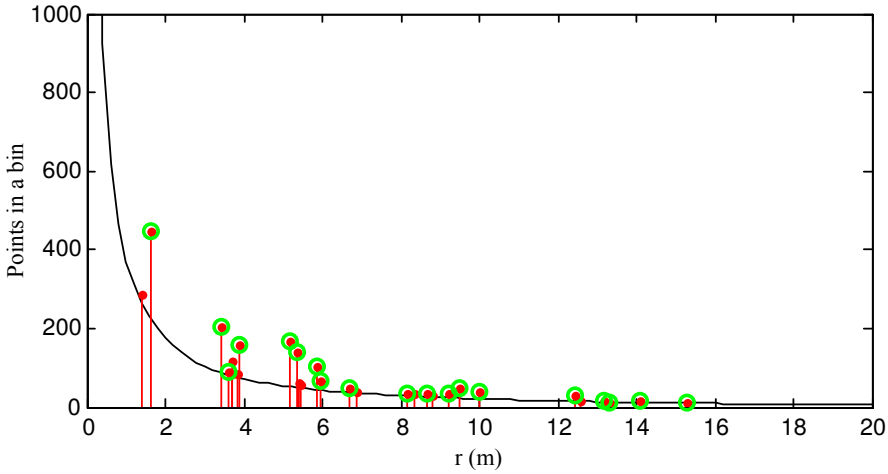
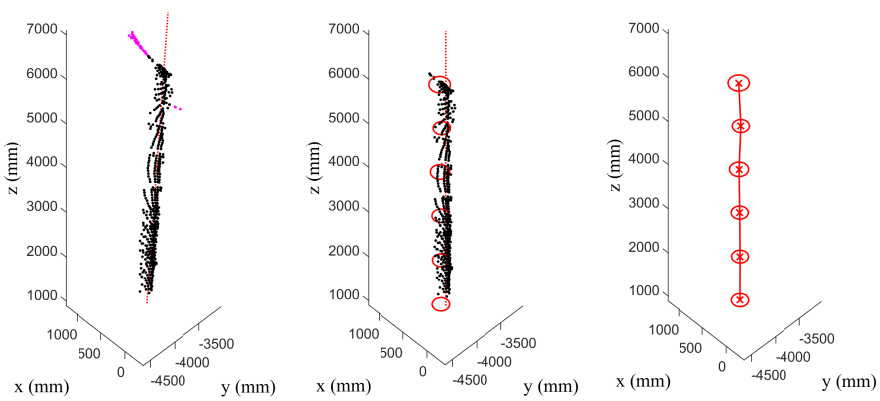


Figure 4.12. A histogram of tree-associated points in a x-y bin as a function of measuring range. The black curve visualizes the threshold function. The final accepted trees are highlighted with a green circle.

bin around the 1 m neighborhood to be used as a seed in the next tree trunk modeling phase.

In tree trunk modeling, all laser range measurements except the ground associated points around a one-meter range from each accepted tree location in the x and y histogram grid are selected. Next, outliers are filtered by iteratively fitting a line to the points, discarding the farthest points out as outliers (magenta points in Figure 4.13a). This filtering is based on the assumption that the main stem is nearly linear, and branches point nearly perpendicularly outwards from the main stem. Thus, the distribution of laser scanned points around the main stem are reduced significantly after



(a) line fitting **(b)** circle fitting **(c)** a tree model

Figure 4.13. A tree modeling process in three steps, where the outliers are filtered out and the line is fitted in (a), overlapping circles are fitted into the remaining point cloud in (b), and the final tree model is visualized in (c).

some distance. This is searched for using a histogram of distances of points from the fitted line. The first distance, where the histogram value falls below a set threshold is decided as the distance limit from the fitted line to filter out branches and foliage around the tree stem. When this is iterated a few times, most branches are removed from the tree stem point cloud.

The remaining filtered point cloud is then rotated such that the fitted line is aligned towards the z axis for further steps (see Figure 4.13b). The tree trunk is then modeled by fitting circles to the points at 1-meter intervals starting from 1 and ending at 10 meters above the estimated ground surface. The point cloud is split into overlapping segments symmetrically around every height. The data regions are set to overlap by half a meter in both directions, as the circle fitting phase requires as much data as possible.

Circles are then iteratively fitted into each of the height segments where there are sufficient data points to fit a circle. The closed form circle fitting procedure presented by Coope (1993) is used to fit the circle to the measurements. It is used iteratively to minimize the effect of outliers by removing the farthest measurements. This iteration is repeated three times. One standard deviation of distance from inside and outside the fitted circle to the circle is used as the limit to detect outliers from the circle fitting process. The resulting circle fit is visualized in Figure 4.13b. The final tree model is then visualized with a wire frame model where the circle center points are connected to each other in Figure 4.13c.

5. Discussion

A forest is an uncontrolled, unstructured environment where sensor readings are uncertain and difficult to interpret. As explained earlier in Section 1.1, the boreal forest floor consists of rocky, shallow-soiled uplands, wetlands, and poorly drained organic soils. Most forests are, at least to some extent, managed, and they consist primarily of conifers (such as spruces or pines), although some broad-leaved species (such as birches) are also common. Because the environment is complex to interpret even in the most managed forests, any autonomous operation requires a robust perception system. As explained earlier in this work, an autonomous (or semi-autonomous) forest machine must, at minimum, be able to measure 1) its own position in the forest, 2) its own inclination to avoid falling, 3) the pose of its crane and tool, 4) trees and their species and other relevant quality-related parameters, and 5) the ground surface around the machine to enable safe navigation.

Moreover, the perception system must cope with adverse weather conditions and imperfect data. The task is challenging, since forest machines operate in a diverse environment in which tree species, size, and appearance, as well as the amount of undergrowth, change drastically over time and space. Furthermore, commonly used GNSS-based navigation sensors lack sufficient positioning accuracy in dense forests, which increases the challenges of locating the robot and combining multiple measurements around the same forest. Moreover, forestry industry budgets are tight. Thus, to allow commercial forest machine manufacturers to add the technology to their products and thus increase the productivity of forest work, neither the sensors nor the technology can be prohibitively expensive.

5.1 Scientific Contribution

The main scientific contribution of the thesis is the proposal of perception systems for autonomous and semi-autonomous forest machinery. These perception systems can be divided into two main parts: a *proprioceptive*

part for sensing the robot's own state and an *exteroceptive* part for sensing the forest around the robot.

Proprioceptive Perception

The proprioceptive aspects of the system enable it to reliably estimate the crane posture, attitude of the vehicle, and orientation of the freely swaying tool. In addition, in a semi-autonomous system, perception of the human operator is classified as perception of the robot's self.

The minimal instrumentation of a flexible forestry crane using a laser scanner in Publication I proved capable of reliably estimating the posture of the crane and the position of the boom tip. The tip position was successfully estimated in spite of obstacles and foliage that occasionally obstructed the line of sight to the crane boom and boom tip mounted targets. In contrast to traditional joint-angle-based measurement systems, the main benefit of the laser-scanner-based crane-posture measurement system is that crane bending can also be observed, since the boom tip position is measured directly instead of computing it through a kinematic chain of joint angles and links that are assumed to be rigid.

In Publication I, crane posture was measured with respect to the machine's main body with RMSEs of 0.14° and 0.40° for the lift and the transfer joint angles, respectively, and 4.0 and 4.3 cm for the extension length and the tip position, respectively. Moreover, in a noise tolerance test (see Figure 5.1), the proposed method (described in Section 3.2) performed well even when visibility was artificially reduced to 0.1 random obstructions per meter for each laser range measurement. This simulates very dense fog or a snowstorm that would obstruct (i.e., reflect back) every

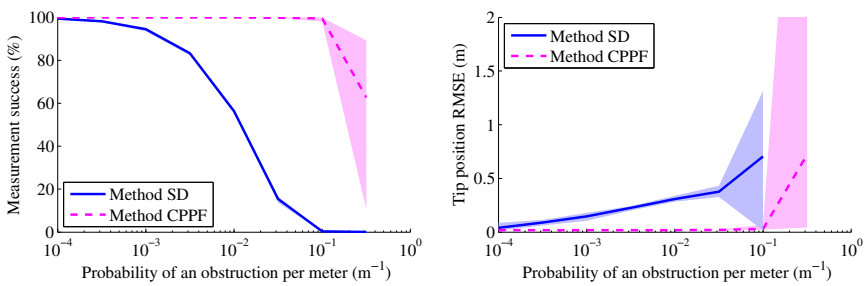


Figure 5.1. In Publication I, the proposed crane posture estimation method (CPPF) was compared to a deterministic target detector method (SD) for tolerance against random obstructions. The left side plot shows the percentage of successful measurements for both methods. The right side plot is a RMSE of the boom tip position. Both measures are drawn as a function of probability of an obstruction (i.e., simulated noise) on the line of sight between the scanner and the actual range measurement. Each noise level was tested 10 times and the area between minimum and maximum values is shaded under each mean curve. Method SD gave zero measurements under the largest amount of noise, and thus those values are missing from the right side plot.

10th laser range measurement during the first meter. Thus, to answer Q_1, the work in Publication I advanced the state of the art of estimating the posture of a forestry crane and the position of the boom tip by providing the first viable laser-scanning-based forest-machine-crane-posture measurement solution for a forest environment.

The attitude estimation method in Publication II showed that it is possible to accurately estimate the attitude and drifts of gyroscopes without relying on other than inertial measurements (e.g., without commonly utilized magnetometer measurements, which are difficult to use in forest machinery). Although MEMS IMU sensors must be calibrated in order to be accurate, as shown in Publication II, in which the calibration was performed as a function of sensor temperature, the sensor drifts develop over time as explained in Section 2.2. Thus, the only option is to calibrate the sensor during use. Then, as explained in the parameter estimation part of Section 2.1, this gyroscope drift can be defined as a parameter and estimated online with attitude in the same filter, as shown in Section 3.3, to provide an estimate of the gyroscope biases combined with the direction of gravity.

In addition to drifting gyroscope biases, the sensor fusion algorithm in Section 3.3 was implemented so as to withstand the random transient non-gravitational acceleration and noisy measurements normally present in any moving vehicle, such as a forest machine. In Publication II, the proposed algorithm was compared to two state-of-the-art open-source algorithms in multiple tests with two different low-cost IMUs and an accurate reference pose. The proposed method transpired to be the most accurate algorithm in a rotation test (RMSEs of 1.57°, 0.56°, and 0.61°, for yaw, pitch, and roll Euler angles, respectively) and also proved highly accurate in a test to measure tolerance against rapid linear accelerations (RMSEs of 1.19°, 0.42°, 0.17° for yaw, pitch, and roll Euler angles, respectively). Furthermore, as shown in Figure 5.2, only the proposed method tolerated large induced biases in gyroscope measurements that could be caused, for instance, by poor calibration or temperature changes in the sensor chip. These properties allowed the low-cost IMU to reliably measure the attitude of a forest machine in the presence of noise and non-gravitational accelerations (Q_2).

In Publication III, the orientation of a freely swaying tool, such as a grapple or a point cleaning tool, was acquired by using two similar IMUs and a sensor fusion algorithm to estimate the orientation of a forest machine tool. The tool was mounted on the boom tip of a forestry crane using a rotator link and a rotator motor to enable free swaying and controlled rotation of the tool. As found in Section 3.4, installing traditional positioning sensors on each of these joints is challenging due to the mechanical structure of the rotator link. Therefore, the EKF for the dual IMU system presented in Section 3.5 offers a practical and affordable way to measure the orientation

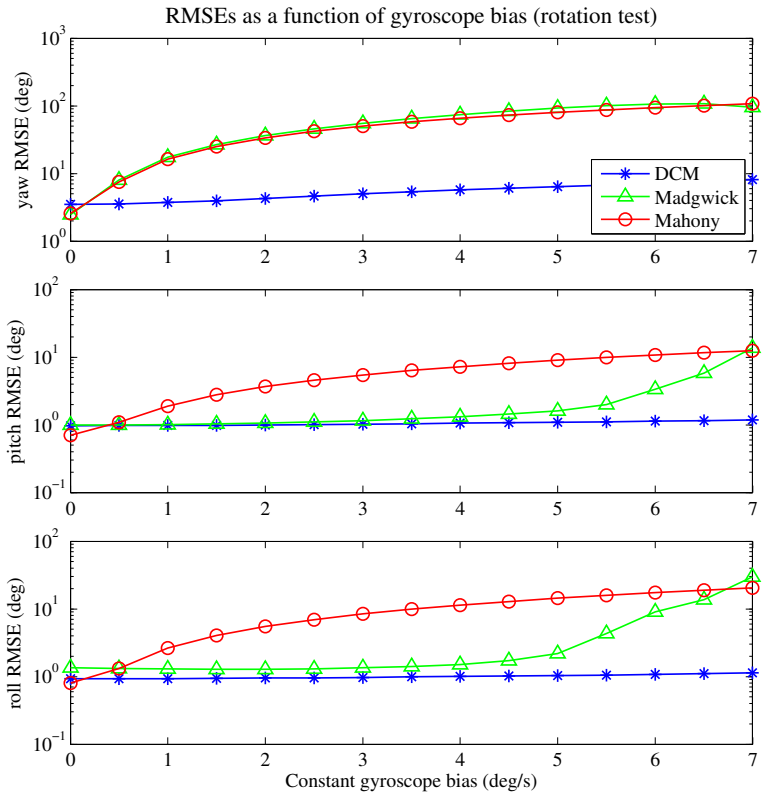


Figure 5.2. In Publication II, the proposed DCM based attitude estimation method was compared to two open-source attitude estimation methods by Madgwick et al. (2011) and Mahony et al. (2008). In a test for tolerance against induced gyroscope bias, the proposed method was the only method capable to estimate the attitude with small RMSE in all tested scenarios.

of the forestry crane tool. To the best of the author's knowledge, no other works than Publication III have proposed a method for measuring the angles of a freely hanging rotator link and the rotator motor. This method is able to estimate the 3D angular position and angular velocities of the freely hanging tool with an accuracy of a few degrees (errors of -5° to 3° for the rotator angle and 1° to 2° for the freely swaying angles). This accuracy is adequate for most practical use cases where tool orientation is required (e.g., picking up logs or cleaning young stands). Thus, to answer Q_3, the work in Publication III enabled the instrumentation of a rotator-link mechanism by estimating the three-dimensional orientation of the tool of the forest machine.

In a semi-autonomous system, a human operator still sits inside the robotized forest machine. When the human operator and the semi-autonomous forest machine work together, it is essential that real-time information is provided to the operator in a way that allows intuitive understanding of what is occurring around the robot. Previous research has hypothesized

that this could be achieved using an augmented reality user interface, but such a solution requires the operator's head pose to be measured with sufficient accuracy and minimal delay. Publication IV demonstrated the robot's ability to perceive the operator's head pose inside the forest machine cabin using a head mounted camera and IMU sensor. As described in Section 3.6, an EKF was used to combine camera and IMU measurements to estimate operator head pose in a forest machine cabin. This provided a real-time estimate of the head pose in order to augment the operator's view with the measurements around the machine. In sum, Publication IV demonstrated that an augmented reality system can be set up in the forest machine cabin to show the forest machine operator laser scanned data around the forest machine in real time (Q_4).

Exteroceptive Perception

In turn, the exteroceptive parts of the system allow it to perceive the 3D structure of the surrounding environment, including tree trunks and the ground surface underneath, to classify species of detected trees and perform real-time visual servoing of the forestry crane to facilitate an automated task such as automatic point cleaning.

As found in Section 2.4, lidars are the best sensors for measuring accurate 3D information from the surrounding forest. From the three different 3D lidars built from lower cost 2D laser scanners (Publications I, VI, and VII), the rotating 45° tilted scanning plane transpired to be the best compromise for point density and FoV. Point density was rather equal for the whole FoV, as shown in Figure 4.4b of Section 4.1. The FoV of the rotating tilted scanner is $360° \times 90°$, leaving only 45° upward and downward cones uncovered. However, if mounted on top of a forest machine cabin, the cabin roof blocks the downward pointing cone. Moreover, the upward pointing cone covers the sky above, from which the measurements are of mostly no interest. As a result, most measurements in this scanning configuration are taken rather equally from the most interesting areas around the forest machine (Q_9).

Although the other scanning configurations were less suitable as the main sensor of an autonomous forest machine, they also exhibited some benefits. The vertical lidar mounted on the side of the forestry crane in Publication I allowed simultaneous measurement of the posture of the crane and the measurement of the working environment. This is a major cost saving, since the same single 2D lidar could be used to measure the crane posture and environment around the vehicle. In turn, the swiveling scanning configuration in Publication VI enabled the collection of a denser point cloud in a narrower FoV, which was beneficial when fusing the lidar and camera data together. This is especially advantageous when the resolution of the equal-density-but-sparse rotating tilted lidar is insufficient

for a tree detection or classification task (Q_9).

Tree trunk and ground modeling was demonstrated in Publication VII by using a lidar built from a rotating tilted 2D laser scanner. The 45° tilted scanning plane enabled the segmentation of ground- and tree-trunk-associated points from individual scan lines. Because of the tilted scanning plane, the primarily horizontal ground surface and vertically growing trees could be detected from individual laser scan profiles. A combination of measurements from the full rotation of the 3D lidar enabled reliable detection of pillar-like tree trunks from the data and the construction of a local grid-type ground model underneath the trees (see Figure 5.3). To answer Q_7, Publication VII introduced a plausible method for a forest-machine-mounted laser scanner to measure the surrounding forest and segment trees and ground surface from the measurements. In Publication VII, tree trunks nearer than 8 meters to the lidar were measured with nearly normally distributed errors less than ±20 mm for the radius. Thus, averaging multiple measurements should significantly increase diameter measurement accuracy.

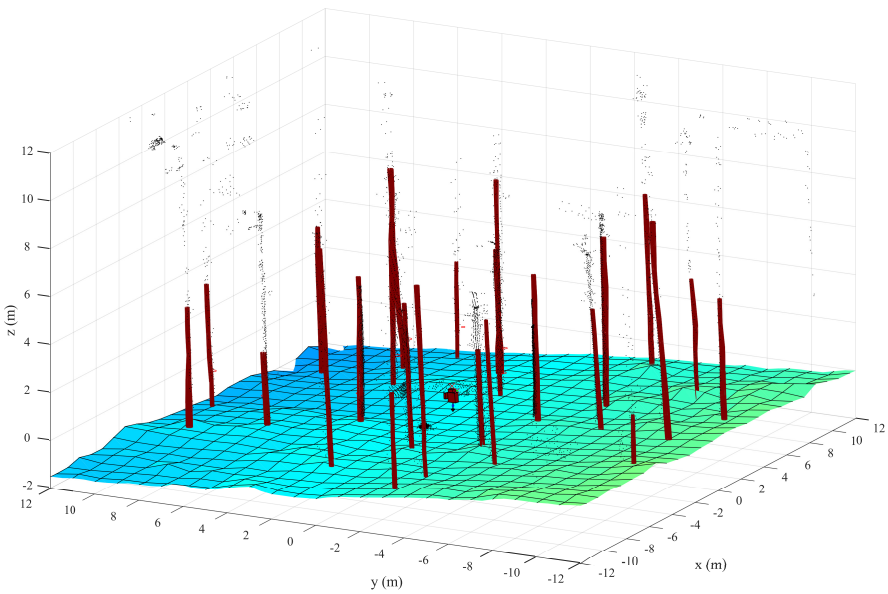


Figure 5.3. In Publication VII, two 45° tilted 2D scanners were rotated to measure and model the ground and tree trunks. In the figure, an example of point cloud collected during 2 seconds (one rotation of the motor in the origin) is modeled as ground using 1 m grid and tree trunks using red cylinders. The small black dots visualize the measured point cloud.

Since a lidar sensor only provides information from one wavelength, and the resolution of the measurements is usually quite low, sensor fusion with a machine vision camera can help significantly in tree species classification. The improvement by adding a camera is twofold. Firstly, the increased resolution enables texture-based recognition, and, secondly, the color chan-

nels of the camera enable the usage of intensity information on multiple wavelengths. In Publication VI, tree species were classified in a young mixed-species forest by using combined lidar and machine vision sensors and a sensor fusion approach to detect, segment, and classify young trees growing next to each other (see Figure 5.4). The work demonstrated that by using a calibrated 3D lidar and a machine vision camera together, it is possible to segment trees based on the range data and then use the color and texture features of the segments to determine tree species (Q_6). The classification result in Publication VI was not perfect, but it should be sufficient to enable the use of automated decision making in an autonomous or semi-autonomous forest machine.

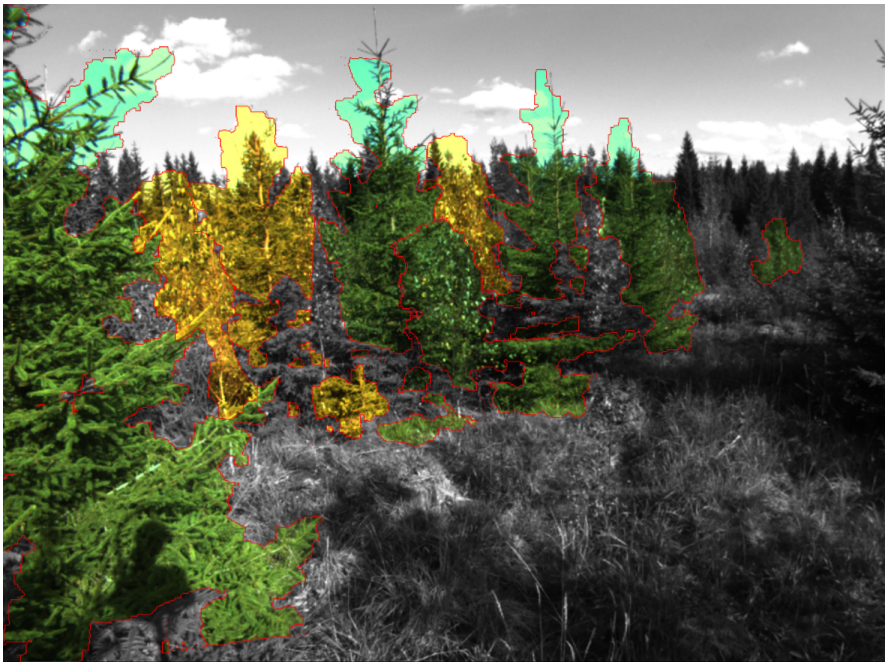
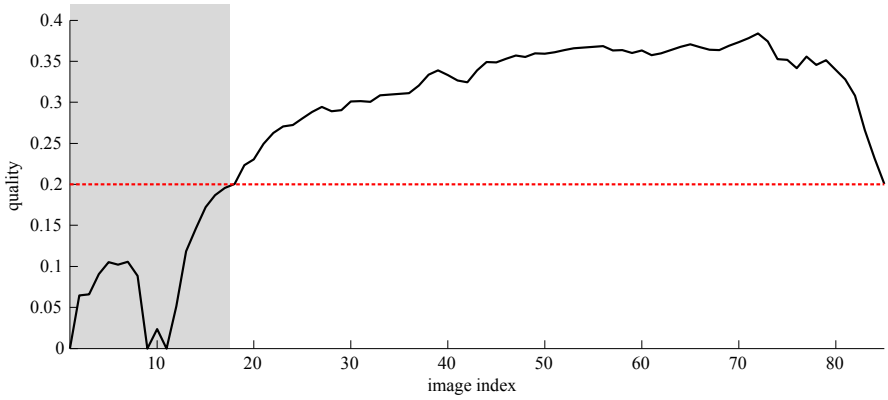
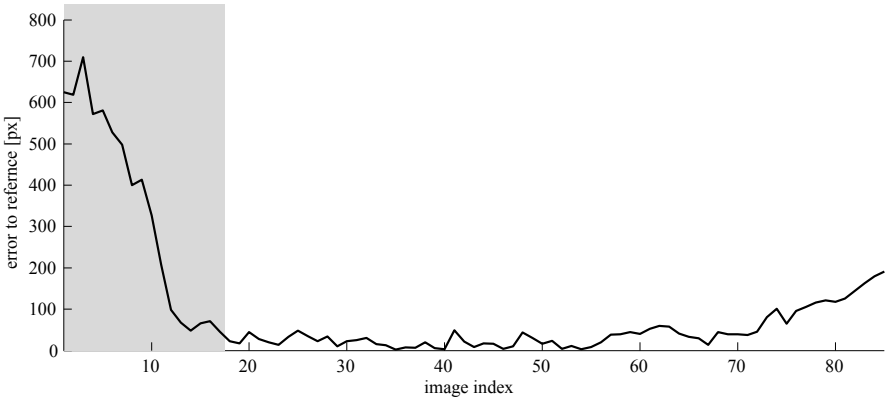


Figure 5.4. In Publication VI, young mixed-species forest was segmented for individual trees and the tree species were classified using a sensor fused 3D lidar and a camera. This figure shows an example segmentation and classification results between spruces (green) and birches (yellow). Note that because of the lidar based segmentation, undergrowth is not segmented nor detected as a tree.

A similar machine vision solution using color and texture information was also used in a demonstration of an autonomous forest machine prototype in Publication V. In the demonstration, a machine vision system was used to detect and classify young spruce trees and help the autonomous forest machine visually servo the point cleaning tool mounted on an autonomous forestry machine to automatically clean vegetation around young spruce trees. The demonstration was performed in the Neosilvix project (Aalto University, 2013) to show that it is possible to use a fully autonomous forest machine in a repetitive but challenging silvicultural task. In the



(a) Estimated detection quality to decide when spruce detection is valid



(b) Detection error against a manually labeled reference spruce positions in the image

Figure 5.5. In Publication V, detection of young spruces was demonstrated in real time. The gray color visualizes the region in which, based on the estimated quality value in (a), the system decided that the detection should not be used. The positioning error in (b) was measured against a manually labeled spruce center points from the recorded images.

demonstration[26], all the young spruce trees were found, and the surrounding competing vegetation was successfully cleaned autonomously (Q_5). Publication V demonstrated the ability of the machine vision algorithm to detect and track a young spruce tree in real-time with an average error of less than 50 pixels for the pose of the tree in a high-resolution camera image of 1280x960 pixels (see Figure 5.5). This accuracy was sufficient for an automated crane controller (Kalmari, 2015) to perform visual servoing using real-time video-based detection as an input.

[26] A video of the demonstration is available at `https://youtu.be/n3xeEsscunA`

5.2 Limitations of the Proposed Solutions

All the solutions proposed in this thesis are compromises caused by computational limitations, requirements for low-cost solutions, and difficult weather conditions, to name but a few limiting factors. This section describes the most important limitations of the methods and sensor solutions.

The crane posture estimation method proposed in Publication I and explained in Section 3.2 is an optical method that is limited in its ability to detect the two targets explained in Section 3.1. In the tests, it was shown to withstand a high amount of noise and occlusions, but if the optical path is, for some reason, obstructed for an extended length of time, the method will evidently fail. For these reasons, traditional joint angle position sensors are more reliable; however, they are more costly than the proposed laser scanner solution, and they are unable to measure the bending of the crane under load.

The laser scanner was also used simultaneously to collect measurements from the environment, as shown in the crane-boom-mounted vertical 2D scanner presented in Section 4.1. However, this instrumentation can only measure distances on one vertical line in the direction of the crane. Nonetheless, this limitation could be easily avoided by using a modern, low-cost, multi-beam lidar, for example, a 16 beam Velodyne Puck Lite (Velodyne Lidar, 2019b), a 32 beam Ouster OS1 (Ouster, 2021), or a 40 beam Hesai Pandar40P (Hesai, 2020), instead of the currently used SICK LMS221 2D scanner (SICK AG, 2006). The other significant drawback is that, in this scanning configuration, a large share of the data is lost towards the ground and sky. However, this could be easily avoided by using a low-cost sensor which provides a narrower FoV such as Neuvition Titan P1 (Neuvition, 2023), which provides one scan line in 135° FoV. This solution may be a cost-effective alternative for a forest robot in cases where a minimal number of sensors are used to simultaneously measure the crane posture and the working environment around the crane in the same coordinate frame.

The attitude estimation method proposed in Publication II and explained in Section 3.3 is robust against transient non-gravitational accelerations, but it contains some limitations. Firstly, it cannot function well in the presence of constant non-gravitational acceleration, which is the case, for example, if the sensor is mounted on a continuously rotating or fast moving object. However, the proposed algorithm is a good choice when carried by a drone, human, or slowly moving vehicle, such as a forest machine.

Secondly, the algorithm cannot estimate the absolute heading angle; rather, it is only capable of calculating the relative and slowly drifting heading change from the initial orientation. The method is designed to work without any sensors other than the triaxial accelerometer and gyroscope, and thus other sensors, such as a compass or magnetometers,

would be required to allow the robot to distinguish the absolute heading angle. However, as forest machinery is mostly built from ferromagnetic metals, uses electric currents to control and power its equipment, and may work in magnetically challenging conditions, such as under power lines, the attitude estimation algorithm is intentionally separated from any dependence on commonly used magnetic sensors. In future research, other methods, such as GNSS or map-based navigation methods, should be integrated with the proposed attitude estimation method to overcome this limitation.

As found in the previous section, the method for estimating IMU-based forest-machine-tool orientation proposed in Publication III and explained in Sections 3.4 and 3.5 enables the practical measurement of the angles of a freely swaying forest machine tool, such as a grapple or a point cleaning tool. The limitation of this method is its accuracy, since it contains errors as large as 5°. The method could most likely be improved in future work by using a more sophisticated model to take account of the accelerations measured in the lower IMU. In the proposed solution, these were omitted to simplify the model to allow for its easy implementation in an embedded system providing real-time estimates.

The head pose estimation method proposed in Publication IV and explained in Section 3.6, which enabled the use of an augmented reality user interface in a forest machine cabin, was able to measure reliably the slow and fast motion of the operator's head in the cabin. However, the cabin also moved, since, in the forest machine used by the study, the cabin was suspended with springs to improve the operator's working ergonomics. The proposed method was unable to measure this cabin motion separately. Since the head pose was measured inside the cabin, cabin motion remained a residual error in the real-time data collected by the forest machine that was augmented on the operator's view. In future work, this could be corrected by also estimating the motion of the spring suspended cabin, for example, by using a separate cabin fixed IMU and the attitude estimation method proposed in Publication II.

The color-and-texture-feature-based algorithm for detecting young spruce proposed in Publication V and explained in Section 4.3 with the hardware-related details shown in Section 4.2 was built with real-time considerations in mind. This meant that real-time-capable operation was the key requirement, and the quality of the detection was subordinate. Similarly, the tests in Publication V were also designed with real-time operation in mind, and young tree detection quality was not the main focus. However, the detection quality was still good enough for detecting all trees reliably. Although the method proved capable of detecting all young spruce in all test cases in the real-time demonstration, currently available state-of-the-art deep learning methods would most likely outperform the spruce detection method in both speed and accuracy (see, e.g., Wäldchen, Rzanny, Seeland,

& Mäder, 2018; Kattenborn, Eichel, & Fassnacht, 2019). However, this part of the work was already published in 2013, when neither deep learning methodology nor the tools to use it in real time had yet matured (see the substantial growth after 2013 in Q. Li et al., 2021). In future work, modern methods to enable real-time detection of young spruces would need to be implemented for the same purpose of visual servoing the crane if similar tasks are attempted.

After learning from the previously noted tree classification challenges in Publication V, a wider search for suitable tree segmentation and classification methods was performed while classifying the trees with the methods surveyed in Publication VI and explained in Section 4.4. The work found many better alternatives for tree species classification, but classification accuracy nevertheless remains the key limitation here. Today, more modern, deep-learning-based solutions would probably outperform the proposed feature-based methods, in which the best classification was a search using the engineered selection of features (see, e.g., Wäldchen et al., 2018; Kattenborn et al., 2019). In the future work, deep-learning solutions should be attempted for segmenting and classifying tree species in a young forest.

The swiveling 2D scanner configuration which was used in Publication V and explained in Section 4.1 was employed to acquire a 3D point cloud around the camera data to perform sensor fusion between the sensors. One of the disadvantages of the swiveling 2D scanner prototype was that the vertical FoV was limited to ±30°. This was slightly too great a limitation in the forest environment, where some of the tops and bottoms of the trees were cropped out. The sensor also suffers from limited FoV in the horizontal direction, as it cannot see anything at the rear of the sensor. The swiveling motion was also a notable limitation, since the system geometry and motion caused periodic accelerating and decelerating sideways movements. This induced many low frequency vibrations in the system, which decreased the accuracy of the point cloud.

Of the scanning configurations tested in this thesis, the rotating tilted scanning configuration proposed in Publication VII and explained in Section 4.1 proved the best compromise. It provides rather equal distribution of the point cloud around the whole space surrounding the vehicle; however, it cannot see anything directly upwards or downwards, which might be a limitation in some tasks. Nonetheless, in forest machinery, if the sensor is mounted vertically, for example on top of a forest machine cabin, the measurements cover the whole environment around the machine rather well. The rotating configuration also possesses limited resolution, and thus it is unsuitable for taking denser, more focused measurements from an area of interest. For this task, the swiveling motion used in Publication VI is more apt.

The proposed tree trunk and ground surface detection methods explained

in Section 4.5 and demonstrated in Publication VII were able to detect tree stems and the ground surface underneath them in a mature pine forest. The method was neither developed nor tested for younger mixed species or spruce forest, in which the tree stems are less visible. For these tasks, a better sensor and more modern algorithms (e.g., E. Hyyppä, Kukko, et al., 2020; E. Hyyppä, Hyyppä, et al., 2020) should be used in future work.

5.3 Discussion on the Selected Sensors and Methods

In this thesis, no single sensor alone is able to provide perception capabilities. Instead, this work proposes that the perception capabilities of an autonomous forest machine are built as a sensor fusion solution. As found in the sensor fusion part of Section 2.1, the calibration of sensors and their mutual configuration is important. In addition, time synchronization of sensors and estimation of the delays in their data streams are vital in order for the same event to be combined between data coming from different sensors. For example, in Publication VI, it was necessary to calibrate the lidar and the camera individually to provide their measurements in a known coordinate frame. Secondly, it was necessary to calibrate their mutual configuration in order to project lidar data in the camera frame. Thirdly, when these sensors are used in motion, their mutual delay in measurements must be known as well as their motion.

The Importance of Inertial Measurements

Since forest machines move relatively slowly on the forest floor, most changes in the lidar and camera data are caused by the rotation of the vehicle and the sensors mounted on it. The attitude information provided by an IMU can be used in many ways. Firstly, it helps balance the robot and prevent it from falling over when traversing uneven terrain and steep slopes. Secondly, when combined with remote sensing sensors, the IMU can reveal the orientation of the vehicle and sensors mounted on it. This simplifies sensor data processing and sensor fusion, since the location where the sensor points is already known.

The attitude estimation algorithm in Publication II was successfully used as the basis for the method in Publication IV, and it provides the important attitude estimate required in Publication III as a part of the boom tip IMU estimate. The inertial sensors in publications V, VI, and VII can be used as a background tool to orientate the optical sensors to make sense of the data if the platform operates in sloped or variable terrain. A robust attitude estimate (as proposed, e.g., in Publication II) is an essential part of any perception system in a forest environment (Q_{10}). The method employed in Publication II has been open sourced in GitHub (Hyyti, 2015) and already

used by some other authors in their research (e.g., Ylikorpi, Halme, & Forsman, 2017; Sandru, Hyyti, Visala, & Kujala, 2020).

As mentioned in Section 3.3, in addition to Publication II, only a small number of algorithms (Hamel & Mahony, 2006; Mahony et al., 2008; Wu et al., 2014) have been proposed to estimate gyroscope biases without any extra sensors in addition to the triaxial accelerometer and gyroscope. The low number of proposed solutions may relate to the difficulty of the challenge. With standard orientation representations, there are at least three unknown parameters representing the rotation in addition to three unknown gyroscope drift parameters. By contrast, there are only five independent measurements, three angular velocities from the gyroscopes, and one vector measurement from the accelerometers, giving two independent measurements and the magnitude of the Earth's gravity. This means that the system is underdetermined, since there are fewer measurements than unknowns. Many other authors have added some extra sensors, such as magnetometers, to the setup to overcome this problem (e.g., Gebre-Egziabher et al., 2004; Lou et al., 2011; Edwan et al., 2011; Madgwick et al., 2011). However, in Publication II the heading angle is decoupled from the attitude estimate by using the direction of gravity instead of the full attitude in the filter, and thus there are only five independent variables estimated in the proposed filter. Therefore, the system is balanced, and it is able estimate the attitude and the gyroscope biases together.

The Role of Optical Measurements

As found in Section 2.4, a color camera is an important tool for measuring texture and color information in a forest. Cameras can be of great benefit for tree and tree species recognition, as shown in publications V and VI. The resolution of commonly used, relatively low-cost machine vision cameras is already high enough to detect and classify the texture of the tree trunk and leaves and separate them from undergrowth, ground, rocks, and other common objects in a forest. In addition, color information is essential in a forest to distinguish between species, as shown in the publications. Hyperspectral sensors could also be great tools for species recognition (e.g., Näsi et al., 2016), but they are somewhat more expensive and thus not considered in this work.

Section 2.4 also shows that a lidar, especially a pulsed-type ToF lidar, is better suited for measuring the 3D structure of the surrounding environment than a radar or sonar, since the angular resolution of lower cost radars is a limiting factor. Sonars, on the other hand, work best at short distances, since the speed of sound in air is a limiting factor when working with distances of more than a few meters. However, radars, especially at shorter millimeter wavelengths, could provide the extra benefit of measuring through forest foliage, revealing the ground surface and tree stems

(Hyyti, 2012). Millimeter-wave radars were previously highly expensive and were thus excluded from this work. Future work should study the capabilities of current low-cost sensors (e.g., Almalioglu, Turan, Lu, Trigoni, & Markham, 2020) in the forest environment. Similar to state-of-the-art radars, commercial 3D lidars were too expensive to be mounted on forest machinery. Thus, various constructions to acquire 3D data using lower cost 2D scanners were attempted in this work. The rotating tilted laser scanner, which proved the best compromise in this research, is still the focus of my current work with rotating multi-beam lidars.

Handling Uncertainty of the Environment

Autonomous forest machines should be able to cope with random errors or faults. As explained in the machine perception part of Section 1.1, probabilistic methods are able to assess the validity of the measurement through the probabilities associated with the estimate. This can be used to detect a bad estimate and trigger a safe process to reinitialize the system. For example, the crane posture estimation method in Section 3.2 was built to fail in a controlled manner. In Publication I, a quality value was computed from the probability mass of the given estimate to indicate the ratio of the probability mass around the estimate. Thus, the quality value was high when most of the particles lay around the current estimate. Conversely, the quality was low when there were multiple hypotheses or when the estimated probability mass had scattered around the available state space, indicating that the system was unaware of the location of the crane. With this single value indicating the system's trust in its own estimate, it was simple to use a hard-coded minimum quality limit to trigger reinitialization.

Similarly, in Publication V, the quality of tree detection was estimated by counting a quality value as the sum of spruce votes given by the k-NN classifier inside a tree-sized circle centered on the estimated location and dividing it by all the spruce votes in the image. The quality value helped the system to determine whether there were any spruce trees at all in the image (see Figure 5.5a). This kind of self-aware estimation of robot states is particularly important in autonomous systems, where the robot must know when it is not measuring something reliably. This single value to indicate trust in the crane posture measurement easily allows the robot to be programmed to stop safely if the quality remains too low for an excessive period of time.

Real-Time Capability

All algorithms intended for autonomous-forest-machine use must work efficiently, and it is also necessary for them to be real-time capable. Real-

time capable computation is essential in perception methods to allow the data to be used as a part of a feedback loop in an autonomous or semi-autonomous forest machine. In Publication I, the computational efficiency of the particle filter algorithm was increased significantly by using an approximate inverse measurement model instead of the conventional (forward) model. The computational benefit is derived from there being just two small identical targets in the environment but 721 laser range observations in each scan. Through inversion, the ray-casting of all laser range observations to a surface defined by the two targets is avoided for each particle. Instead, only the fit of the target needs to be tested in every laser scan measurement, which is efficient and fast (Q_8).

To increase computational efficiency, in publications II, III and IV, the sensor fusion algorithm was implemented in an extended Kalman filter (EKF), which is an efficient probabilistic estimation method requiring significantly less computational resources than, for instance, an unscented Kalman filter (UKF) or particle filter (PF), as found in the nonlinear filtering part of Section 2.1 and visualized in Figure 2.1. All methods in publications II, III, and IV are real-time capable and implemented in C/C++ (Q_8).

Image processing is usually computationally intensive. To expedite this process, in Publication V, the young spruce detection method was implemented with graphics-card-accelerated computing using NVidia CUDA (i.e., Compute Unified Device Architecture) C code (Nvidia, 2012). Since it was necessary for feature extraction and feature classification to work in real time to enable the autonomous forest machine to control the crane based on the camera image stream, it was necessary to optimize multiple aspects of the algorithm to gain more computational efficiency. To enable fast computation, texture feature extraction and classification algorithms needed to be simple enough to be implemented in CUDA code and to be run on a laptop computer with a Core i7-3820QM processor and NVIDIA Quadro K2000M[27]. Since algorithm implementation on the GPU requires significantly more effort than on a CPU, it unreasonable to use the GPU for every algorithm Instead, only the most expensive algorithms were implemented on the GPU (Q_8). These were the Radon transform, the wavelet transform, the feature extraction, and the k-NN classification (see Section 4.3 for algorithm details).

Since the main objective of this thesis is increase the productivity of forest work, good solutions should also be cost effective enough for use in commercial forest machinery. To meet these requirements, the cost of the sensors proposed here is low. For example, the IMUs used in publications II, III, and IV were sensor chips made for cellphone use costing only a few euros per piece. The machine vision cameras and laser scanners, on the

[27] This was a high-end laptop computer at the beginning of the 2010s, when the method was used for Publication V.

other hand, cost a few thousand euros each, and the software was run on normal laptop computers using open-source software. Only algorithm development was performed in the Matlab environment, and all algorithms were implemented in C/C++ code to run them efficiently.

6. Conclusion

In order for an autonomous or a semi-autonomous forest machine to work effectively in a forest, it requires sensors and algorithms that enable it to perceive itself and the surrounding environment. This thesis proposed perception systems to increase the ability of forest machinery to perform some of the work in the forest more autonomously and to increase the productivity of that work. As stated earlier, these perception systems can be divided into two parts: a *proprioceptive* part for perceiving the robot's own state and an *exteroceptive* part for perceiving the forest around the machine.

The proprioceptive part includes 1) a crane posture measurement system (Publication I), 2) an attitude estimation algorithm (Publication II), 3) a tool orientation estimation method (Publication III), and 4) methods to perceive the operator in the cabin and to enable intuitive cooperation between the robot and the human operator in semi-autonomous operations. The methods proposed in Publication IV enable the use of an augmented-reality user interface in a forest machine cabin.

The exteroceptive part includes 1) a 3D lidar capable of measuring a full 360° FoV around the machine to reveal the 3D structure of the surrounding environment (Publication VII), 2) the sensor fusion of the lidar and a machine vision camera to allow tree species detection and classification (Publication VI), and 3) a real-time-capable machine vision system to allow visual servoing of the forestry crane in an autonomous point cleaning operation (Publication V).

The publications included in this thesis show that the level of forest machine automation can be increased by advancing the perception capabilities of the machine. As demonstrated with the prototype implementations and tests in a boreal forest environment, the proposed perception systems enable the forest machine to 1) measure the posture of its crane and the boom tip position within a few centimeters, even under severe disturbances (Publication I), 2) estimate the attitude of the machine within a few degrees, even when the sensor is mounted on the forest machine, causing vibrations and transient non-gravitational accelerations (Publication II), 3)

provide a practical means of measuring the angles of an freely swaying forest machine tool (Publication III), 4) enable augmented reality in a forest machine cabin (Publication IV), 5) perform autonomous point cleaning of young stands by using color-camera-based real-time control (Publication V), 6) detect and classify tree species by using sensor fusion between a camera and lidar (Publication VI), and 7) measure and model tree trunks and the ground surface around the vehicle using a rotating lidar (Publication VII).

To facilitate the adaptation of the proposed methods for industrial purposes, the focus of this thesis was 1) low-cost sensors that could 2) function in a forest environment and 3) withstand harsh weather conditions, such as rain, snow, ice, and direct sunshine as well as a wide temperature range. Moreover, 4) the methods were implemented in a prototype forest machine, and 5) the proposed solutions were tested with a real machine operating in a forest. This approach made the work significantly more challenging, but it ultimately led to the development of several practical innovations, such as lidar-based crane posture measurement (Publication I), IMU-based orientation estimation of a freely swaying forest machine tool (Publication III), and a rotating 45° tilted laser scanning configuration for use in a forest machine (Publication VII).

This thesis answered to ten research questions: Q_1) It demonstrated that a laser scanner mounted on the forestry crane can be reliably used to estimate the posture of the crane and the position of the boom tip; Q_2) it showed that a low-cost IMU can reliably measure the attitude of a forest machine in the presence of noise and non-gravitational accelerations; Q_3) it presented a novel IMU-based instrumentation for a rotator-link mechanism to measure the three-dimensional orientation of the tool of the forest machine; Q_4) it showed that it is plausible to present sensory information in real-time from the perception system to the machine operator using augmented reality in the forest machine cabin; Q_5) it demonstrated autonomous point-cleaning in a young spruce forest; Q_6) it showed tree detection and species classification in a young mixed-species forest; Q_7) it showed how a forest-machine-mounted laser scanner can be used to measure and model tree trunks and ground surface; Q_8) it explained the trade-offs that were implemented to make the proposed methods real-time capable; Q_9) it showed the benefits and limitations of three different actuated laser scanning configurations; and Q_{10}) it explained the importance of inertial measurements in machine perception for forest machinery.

One of the key findings of the thesis is the importance of IMU-based measurements in forest machinery. IMUs are beneficial for 1) revealing the attitude of the machine, 2) estimating the tool orientation, 3) tracking the operator's head pose, and 4) defining the orientation of the other sensors mounted on the machine. In addition, IMU with a sensor fusion algorithm (Publication II) provides an estimate of the non-gravitational acceleration of the part which the sensor is mounted on. The second important finding

is the best scanning configuration for collecting a nearly equally dense 3D point cloud around the forest machine in all relevant directions. This was achieved using the rotating 45° tilted scanning plane in Publication VII. Furthermore, the configuration enabled the detection of ground and tree trunks from individual laser profiles, since trees generally grow vertically and the ground underneath usually lies horizontally.

The third important finding is the usability of color and texture information in forest machinery. Since a lidar sensor only provides information from one wavelength and the resolution of lidar measurements is usually low, sensor fusion with a machine vision camera can help significantly in tree species classification. The improvement by adding a camera is twofold. Firstly, the increased resolution enables texture-based recognition, and, secondly, the color channels of the camera enable the usage of intensity information on multiple wavelengths. As shown in Publication V, a color camera can be used to detect young spruce trees among other competing vegetation in a point cleaning operation. Moreover, as shown in Publication VI, combining the camera with a 3D lidar may enable the forest machine both to detect and segment young trees growing next to each other and to define their species.

This thesis demonstrates the potential of the sensors it proposes to assist the operator, provide accurate forest inventory data, and thus increase the efficiency of forest operations. Section 1.1 showed that better forest data could increase the profitability of the forestry sector by more than 250 million euros a year in Finland alone (Kangas et al., 2019). As there are about 5400 forest machines in Finland (Metsätrans, 2017, 2019; Johnsen, 2019), and if all the reported benefits of better forest data could be gained through forest-machine-mounted sensors and purposely developed probabilistic machine perception methods, each forest machine could potentially produce savings of about 45,000 € a year. Furthermore, all other potential benefits of the added sensors and automation, such as increased productivity and the easier use of machines, should be added to those savings. This will hopefully provide sufficient motivation for the broader use of these methods in the future.

References

Aalto University. (2013). *NeoSilvix project*. https://wiki.aalto.fi/display/autonomous/NeoSilvix+project. Aalto University, Autonomous Systems.

Aegis Electronic Group. (2006). *Foculus FO442B/C IEEE1394 Digital CCD Camera*. http://www.aegis-elec.com/pdf/product/Foculus-FO442BC.pdf. Aegis Electronic Group.

Aggarwal, P. (2010). *MEMS-based integrated navigation*. Artech House.

Ahtiainen, J., Stoyanov, T., & Saarinen, J. (2017). Normal distributions transform traversability maps: LIDAR-only approach for traversability mapping in outdoor environments. *Journal of Field Robotics*, *34*(3), 600–621.

Ali, W., Georgsson, F., & Hellstrom, T. (2008). Visual tree detection for autonomous navigation in forest environment. In *Intelligent vehicles symposium* (pp. 560–565).

Almalioglu, Y., Turan, M., Lu, C. X., Trigoni, N., & Markham, A. (2020). Milli-RIO: Ego-motion estimation with low-cost millimetre-wave radar. *IEEE Sensors Journal*, *21*(3), 3314–3323.

Analog Devices. (2012). *ADXL345* [Datasheet]. Retrieved from https://www.analog.com/media/en/technical-documentation/data-sheets/adxl345.pdf (Rev. G)

Anderson, J. W., & Clayton, G. M. (2014). Lissajous-like scan pattern for a gimballed LIDAR. In *IEEE/ASME international conference on advanced intelligent mechatronics* (pp. 1171–1176).

Argenti, F., Lapini, A., Bianchi, T., & Alparone, L. (2013). A tutorial on speckle reduction in synthetic aperture radar images. *IEEE Geoscience and remote sensing magazine*, *1*(3), 6–35.

Arulampalam, M. S., Maskell, S., Gordon, N., & Clapp, T. (2002). A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Transactions on signal processing*, *50*(2), 174–188. doi: 10.1049/ic:20010246

Asikainen, A., Anttila, P., Verkerk, H., Diaz, O., & Röser, D. (2011). Development of forest machinery and labour in the EU in 2010-2030. In *Proceedings of FORMEC 44th international symposium on forestry*

mechanisation: Pushing the boundaries with research and innovation in forest engineering (p. 8).

Badcock, D. R., Palmisano, S., & May, J. G. (2015). Vision and virtual environments. In K. S. Hale & K. M. Stanney (Eds.), *Handbook of virtual environments: Design, implementation, and applications* (p. 39-85). CRC Press. doi: 10.1201/b17360

Bailey, T., & Nebot, E. (2001). Localisation in large-scale environments. *Robotics and autonomous systems*, *37*(4), 261–281.

Baldwin, G., Mahony, R., Trumpf, J., Hamel, T., & Cheviron, T. (2007). Complementary filter design on the special euclidean group SE (3). In *European control conference (ECC)* (pp. 3763–3770).

Barnea, S., & Filin, S. (2008). Keypoint based autonomous registration of terrestrial laser point-clouds. *Journal of Photogrammetry and Remote Sensing*, *63*(1), 19–35.

Bar-Shalom, Y., Li, X. R., & Kirubarajan, T. (2001). *Estimation with applications to tracking and navigation: theory, algorthims and software*. John Wiley & Sons.

Barshan, B., & Durrant-Whyte, H. F. (1995). Inertial navigation systems for mobile robots. *Transactions on Robotics and Automation*, *11*(3), 328–342.

Bayer, B. E. (1976, 7 20). *Color imaging array.* Google Patents. (US Patent 3,971,065)

Bégin, M.-A., & Hunter, I. (2022). Auto-exposure algorithm for enhanced mobile robot localization in challenging light conditions. *Sensors*, *22*(3), 835.

Behroozpour, B., Sandborn, P. A., Wu, M. C., & Boser, B. E. (2017). Lidar system architectures and circuits. *IEEE Communications Magazine*, *55*(10), 135–142.

Bessière, P., Laugier, C., & Siegwart, R. (2008). *Probabilistic reasoning and decision making in sensory-motor systems*. Springer Berlin Heidelberg.

Bharucha-Reid, A. (2012). *Elements of the theory of Markov processes and their applications*. Dover Publications.

Biber, P., & Straßer, W. (2003). The normal distributions transform: A new approach to laser scan matching. In *IEEE/RSJ international conference on intelligent robots and systems (IROS)* (Vol. 3, pp. 2743–2748).

Bistrov, V. (2012). Performance analysis of alignment process of MEMS IMU. *International Journal of Navigation and Observation*, *2012*, 11.

Blanco Claraco, J. L., & Fernández-Madrigal, J.-A. (2012). *Simultaneous localization and mapping for mobile robots: Introduction and methods: Introduction and methods* (J.-A. Fernández-Madrigal, Ed.). Information Science Reference.

Blocker, M., Bromley, K., & Murdoch, I. (2016). *Global forest, paper, and packaging industry survey: 2016 edition–survey of 2015 results.* Pricewaterhouse Coopers LLP.

Bouguet, J.-Y. (2004). *Camera calibration toolbox for matlab.* Online. Retrieved from http://www.vision.caltech.edu/bouguetj/calib_doc

Burton, P. J., Messier, C., Weetman, G. F., Prepas, E. E., Adamowicz, W. L., & Tittler, R. (2003). The current state of boreal forestry and the drive for change. In P. J. Burton, C. Messier, D. W. Smith, & W. L. Adamowicz (Eds.), *Towards sustainable management of the boreal forest* (pp. 1–40). NRC Research Press, Ottawa, Ontario, Canada.

Campbell, S., O'Mahony, N., Krpalcova, L., Riordan, D., Walsh, J., Murphy, A., & Ryan, C. (2018). Sensor technology in autonomous vehicles: A review. In *29th Irish signals and systems conference (ISSC)* (pp. 1–4).

Campos, M., Litkey, P., Wang, Y., Chen, Y., Hyyti, H., Hyyppä, J., & Puttonen, E. (2020). A terrestrial laser scanning measurement station to monitor long-term structural dynamics in a boreal forest. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, *43*, 27–31.

Candy, J. V. (2007). Bootstrap particle filtering. *IEEE Signal Processing Magazine*, *24*(4), 73–85. doi: 10.1109/MSP.2007.4286566

Carpentier, M., Giguere, P., & Gaudreault, J. (2018). Tree species identification from bark images using convolutional neural networks. In *2018 IEEE/RSJ international conference on intelligent robots and systems (IROS)* (pp. 1075–1081).

Censi, A. (2008). An ICP variant using a point-to-line metric. In *2008 IEEE international conference on robotics and automation* (pp. 19–25).

Chaudhuri, B., Sarkar, N., & Kundu, P. (1993). Improved fractal geometry based texture segmentation technique. *IEE Proceedings E (Computers and Digital Techniques)*, *140*(5), 233–242.

Chaudhuri, S., & Bhardwaj, A. (2018). *Kinesthetic perception: A machine learning approach*. Springer Singapore. doi: 10.1007/978-981-10-6692-4

Chen, S. W., Nardari, G. V., Lee, E. S., Qu, C., Liu, X., Romero, R. A. F., & Kumar, V. (2020). SLOAM: Semantic lidar odometry and mapping for forest inventory. *IEEE Robotics and Automation Letters*, *5*(2), 612–619.

Chen, Y. Q., Nixon, M. S., & Thomas, D. W. (1995). Statistical geometrical features for texture classification. *Pattern recognition*, *28*(4), 537–552.

Cheng, H.-D., Jiang, X. H., Sun, Y., & Wang, J. (2001). Color image segmentation: advances and prospects. *Pattern recognition*, *34*(12), 2259–2281.

Cheng, P., Oelmann, B., & Linnarsson, F. (2011). A local positioning system

for loader cranes based on wireless sensors–a feasibility study. *IEEE Transactions on Instrumentation and Measurement*, *60*(8), 2881–2893. doi: 10.1109/TIM.2011.2121191

Chien, H.-J., Chuang, C.-C., Chen, C.-Y., & Klette, R. (2016). When to use what feature? SIFT, SURF, ORB, or A-KAZE features for monocular visual odometry. In *2016 international conference on image and vision computing New Zealand (IVCNZ)* (pp. 1–6).

Coe, G. (2001). *SRF08 ultra sonic range finder* [Technical Specification]. Retrieved from `https://www.robot-electronics.co.uk/htm/srf08tech.html`

Coope, I. D. (1993). Circle fitting by linear and nonlinear least squares. *Journal of Optimization theory and applications*, *76*(2), 381–388.

Corke, P., Roberts, J., & Winstanley, G. (1998). Vision-based control for mining automation. *IEEE Robotics & Automation Magazine*, *5*(4), 44–49. doi: 10.1109/100.740465

Council of European Union. (2018). Directive (EU) 2018/2001 of the European Parliament and of the Council of 11 December 2018 on the promotion of the use of energy from renewable sources. *Official Journal of the European Union*, *L 328*, 82–209.

Crassidis, J. L., Markley, F. L., & Cheng, Y. (2007). Survey of nonlinear attitude estimation methods. *Journal of guidance, control, and dynamics*, *30*(1), 12–28.

D'Agostini, G. (2003). *Bayesian reasoning in data analysis: A critical introduction*. World Scientific.

Daugman, J. G. (1980). Two-dimensional spectral analysis of cortical receptive field profiles. *Vision research*, *20*(10), 847–856.

Davidson-Pilon, C. (2015). *Bayesian methods for hackers: Probabilistic programming and Bayesian inference*. Pearson Education.

De Luca, A., & Panzieri, S. (1994). An iterative scheme for learning gravity compensation in flexible robot arms. *Automatica*, *30*(6), 993–1002. doi: 10.1016/0005-1098(94)90192-9

Diebel, J. (2006). Representing attitude: Euler angles, unit quaternions, and rotation vectors. *Matrix*, *58*(15-16), 1–35.

Ding, W., Wang, J., Li, Y., Mumford, P., & Rizos, C. (2008). Time synchronization error and calibration in integrated GPS/INS systems. *ETRI journal*, *30*(1), 59–67.

Doer, C., & Trommer, G. F. (2020). An EKF based approach to radar inertial odometry. In *2020 IEEE international conference on multisensor fusion and integration for intelligent systems (MFI)* (pp. 152–159).

Dold, C., & Brenner, C. (2006). Registration of terrestrial laser scanning data using planar patches and image data. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, *36*(5), 78–83.

Donner-Amnell, J., Lehtinen, A. A., & Sæther, B. (2017). Comparing the forest regimes in the conifer north. In A. A. Lehtinen, J. Donner-Amnell,

& B. Sæther (Eds.), *Politics of forests: Northern forest-industrial regimes in the age of globalization* (pp. 255–284). Routledge.

Doya, K., Ishii, S., Pouget, A., & Rao, R. P. N. (2007). *Bayesian brain: Probabilistic approaches to neural coding*. MIT Press.

Duarte, L., Vala, M., Ribeiro, C., Alves, L. N., Fernandes, C. A., & Caldeirinha, R. F. (2022). A feasibility study on real-time concealed object detection in foliage using STDCC radar. In *13th international symposium on communication systems, networks and digital signal processing (CSNDSP)* (pp. 320–325).

Dudek, G., & Jenkin, M. (2010). *Computational principles of mobile robotics*. Cambridge University Press.

Dunbabin, M., & Corke, P. (2006). Autonomous excavation using a rope shovel. *Journal of Field Robotics*, *23*(6-7), 379–394. doi: 10.1002/rob.20132

Edwan, E., Zhang, J., Zhou, J., & Loffeld, O. (2011). Reduced DCM based attitude estimation using low-cost IMU and magnetometer triad. In *8th workshop on positioning navigation and communication (wpnc)* (pp. 1–6).

Edwards, J. (2013). Next-generation robots offer sophisticated mobility, manipulation, and sensing capabilities. *IEEE Signal Processing Magazine*, *30*(5), 11–13.

Elfes, A. (1989). Using occupancy grids for mobile robot perception and navigation. *Computer*, *22*(6), 46–57. doi: 10.1109/2.30720

Elmenreich, W. (2002). *Sensor fusion in time-triggered systems* (doctoral dissertation). Technische Universität Wien, Institut für Technische Informatik, Treitlstr. 3/3/182-1, 1040 Vienna, Austria.

Ercan, Z., Sezer, V., Heceoglu, H., Dikilitas, C., Gokasan, M., Mugan, A., & Bogosyan, S. (2011). Multi-sensor data fusion of DCM based orientation estimation for land vehicles. In *International conference on mechatronics (ICM)* (pp. 672–677).

Errera, M., Dias, T. d. C., Maya, D., & Lora, E. (2023). Global bioenergy potentials projections for 2050. *Biomass and Bioenergy*, *170*, 106721.

Ersson, B. T. (2014). *Concepts for mechanized tree planting in southern Sweden* (doctoral dissertation). Umeå University.

Euston, M., Coote, P., Mahony, R., Kim, J., & Hamel, T. (2008). A complementary filter for attitude estimation of a fixed-wing UAV. In *International conference on intelligent robots and systems (IROS)* (pp. 340–345).

Fahlvik, N., Ekö, P. M., & Petersson, N. (2015). Effects of precommercial thinning strategies on stand structure and growth in a mixed even-aged stand of Scots pine, Norway spruce and birch in southern Sweden. *Silva Fennica*, *49*(3), 17.

Feng, C., Dong, S., Lundeen, K. M., Xiao, Y., & Kamat, V. R. (2015). Vision-based articulated machine pose estimation for excavation monitor-

ing and guidance. In *International symposium on automation and robotics in construction* (Vol. 32, p. 1).

Ferreira, J., Lobo, J., Bessiere, P., Castelo-Branco, M., & Dias, J. (2013). A Bayesian framework for active artificial perception. *IEEE transactions on cybernetics*, *43*(2), 699–711.

Ferreira, J. F., & Dias, J. M. (2013). *Probabilistic approaches to robotic perception*. Springer International Publishing.

FIBIC. (2014). *Intensive forestry – more efficient production of forest-based raw materials*. Online. Retrieved 2017-03-08, from `http://clicinnovation.fi/intensive-forestry-more-efficient-production-of-forest-based-raw-materials`

Fix, E., & Hodges, J. L. (1989). Discriminatory analysis. nonparametric discrimination: Consistency properties. *International Statistical Review/Revue Internationale de Statistique*, *57*(3), 238–247.

Forster, C., Pizzoli, M., & Scaramuzza, D. (2014). SVO: Fast semi-direct monocular visual odometry. In *IEEE international conference on robotics and automation (ICRA)* (pp. 15–22).

Forsyth, D. (2017). *Probability and statistics for computer science*. Springer International Publishing.

Fourier, J.-B. J. (1822). *Théorie analytique de la chaleur*. Firmin Didot.

Fowler, K. R. (2005). Transient response for automatic gain control with multiple intensity thresholds for image-intensified camera. *IEEE transactions on instrumentation and measurement*, *54*(5), 1926–1933.

Fox, D., Burgard, W., & Thrun, S. (1999). Markov localization for mobile robots in dynamic environments. *Journal of artificial intelligence research*, *11*, 391–427.

Foxlin, E. (1996). Inertial head-tracker sensor fusion by a complementary separate-bias Kalman filter. In *Proceedings of the virtual reality annual international symposium* (pp. 185–194).

Fu, L., Gao, F., Wu, J., Li, R., Karkee, M., & Zhang, Q. (2020). Application of consumer RGB-D cameras for fruit detection and localization in field: A critical review. *Computers and Electronics in Agriculture*, *177*, 105687.

Gabor, D. (1946). Theory of communication. part 1: The analysis of information. *Journal of the Institution of Electrical Engineers-Part III: Radio and Communication Engineering*, *93*(26), 429–441.

Garrido-Jurado, S., Muñoz-Salinas, R., Madrid-Cuevas, F. J., & Marín-Jiménez, M. J. (2014). Automatic generation and detection of highly reliable fiducial markers under occlusion. *Pattern Recognition*, *47*(6), 2280–2292. doi: 10.1016/j.patcog.2014.01.005

Gauss, C. F. (1809). *Theoria motus corporum coelestium in sectionibus conicis solem ambientium*. sumtibus Frid. Perthes et I.H. Besser.

Gauss, C. F. (1823). *Theoria combinationis observationum erroribus minimis obnoxiae*. Apud Henricum Dieterich.

Gebre-Egziabher, D., Hayward, R. C., & Powell, J. D. (2004). Design of multi-sensor attitude determination systems. *Transactions on aerospace and electronic systems*, *40*(2), 627–649.

Gellerstedt, S. (2002). Operation of the single-grip harvester: motor-sensory and cognitive work. *International journal of forest engineering*, *13*(2), 35–47.

Gellerstedt, S., & Dahlin, B. (1999). Cut-to-length: The next decade. *Journal of Forest Engineering*, *10*(2), 17–24.

Girshick, A. R., Landy, M. S., & Simoncelli, E. P. (2011). Cardinal rules: visual orientation perception reflects knowledge of environmental statistics. *Nature neuroscience*, *14*(7), 926.

Gordon, N. J., Salmond, D. J., & Smith, A. F. M. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. In *IEE proceedings F (radar and signal processing)* (Vol. 140, pp. 107–113). doi: 10.1049/ip-f-2.1993.0015

Granshaw, S. I., & Fraser, C. S. (2015). Computer vision and photogrammetry: Interaction or introspection? *The Photogrammetric Record*, *30*(149), 3–7.

Gustafsson, F., Gunnarsson, F., Bergman, N., Forssell, U., Jansson, J., Karlsson, R., & Nordlund, P.-J. (2002). Particle filters for positioning, navigation, and tracking. *IEEE Transactions on signal processing*, *50*(2), 425–437. doi: 10.1109/78.978396

Haar, A. (1910). On the theory of orthogonal function systems. *Mathematische Annalen*, *69*(3), 331–371.

Häggström, C. (2015). *Human factors in mechanized cut-to-length forest operations* (doctoral dissertation). Umeå University.

Hague, T., Marchant, J. A., & Tillett, N. D. (2000). Ground based sensing systems for autonomous agricultural vehicles. *Computers and Electronics in Agriculture*, *25*(1-2), 11–28.

Hakkila, P. (1989). Logging in Finland. *Acta Forestalia Fennica*, *207*, 1–39.

Hakobyan, G., & Yang, B. (2019). High-performance automotive radar: A review of signal processing algorithms and modulation schemes. *IEEE Signal Processing Magazine*, *36*(5), 32–44.

Halmetschlager-Funek, G., Suchi, M., Kampel, M., & Vincze, M. (2018). An empirical evaluation of ten depth cameras: Bias, precision, lateral noise, different lighting conditions and materials, and multiple sensor setups in indoor environments. *IEEE Robotics & Automation Magazine*, *26*(1), 67–77.

Hämäläinen, J., Strandström, M., Saarinen, V.-M., Hynynen, J., Saksa, T., & Hyyti, H. (2013). Koneellisen taimikonhoidon kustannustehokkuuden parantaminen (improving the cost effectiveness of mechanized tending). *Metsätehon raportti*, *228*, 27. (in Finnish)

Hamel, T., & Mahony, R. (2006). Attitude estimation on SO[3] based on

direct inertial measurements. In *Proceedings of 2006 IEEE international conference on robotics and automation (ICRA)* (pp. 2170–2175).

Hanning, T. (2011). *High precision camera calibration*. Springer.

Hansen, A. C., Clarke, N., & Hegnes, A. W. (2021). Managing sustainability risks of bioenergy in four nordic countries. *Energy, Sustainability and Society*, *11*(1), 20.

Hansen, P., Alismail, H., Rander, P., & Browning, B. (2013). Pipe mapping with monocular fisheye imagery. In *IEEE/RSJ international conference on intelligent robots and systems* (pp. 5180–5185).

Hansen, R. E. (2013). Synthetic aperture sonar technology review. *Marine Technology Society Journal*, *47*(5), 117–127.

Hansson, A., & Servin, M. (2010). Semi-autonomous shared control of large-scale manipulator arms. *Control Engineering Practice*, *18*(9), 1069–1076.

Haralick, R. M., Shanmugam, K., & Dinstein, I. (1973). Textural features for image classification. *IEEE Transactions on systems, man, and cybernetics*, *SMC-3*(6), 610–621.

Hasch, J., Topak, E., Schnabel, R., Zwick, T., Weigel, R., & Waldschmidt, C. (2012). Millimeter-wave technology for automotive radar sensors in the 77 GHz frequency band. *IEEE Transactions on Microwave Theory and Techniques*, *60*(3), 845–860.

Hauglin, M., Hansen, E. H., Næsset, E., Busterud, B. E., Gjevestad, J. G. O., & Gobakken, T. (2017). Accurate single-tree positions from a harvester: A test of two global satellite-based positioning systems. *Scandinavian Journal of Forest Research*, *32*(8), 774–781.

Hayes, M. P., & Gough, P. T. (2009). Synthetic aperture sonar: A review of current status. *IEEE Journal of Oceanic Engineering*, *34*(3), 207–224.

Hellström, T., Lärkeryd, P., Nordfjell, T., & Ringdahl, O. (2009). Autonomous forest vehicles: Historic, envisioned, and state-of-the-art. *International Journal of Forest Engineering*, *20*(1), 31–38.

Hellström, T., & Ostovar, A. (2014). Detection of trees based on quality guided image segmentation. In *Second international conference on robotics and associated high-technologies and equipment for agriculture and forestry (RHEA-2014)* (pp. 531–540).

Hesai. (2020). *Pandar40P 40-channel mechanical lidar* [User Manual]. (402-en-2002A2)

Hillier, N., Ryde, J., & Widzyk-Capehart, E. (2015). Comparison of scanning laser range-finders and millimeter-wave radar for creating a digital terrain map. In *Machine vision and mechatronics in practice* (pp. 23–38). Springer.

Hofmann-Wellenhof, B., Lichtenegger, H., & Wasle, E. (2007). *GNSS–global navigation satellite systems: GPS, GLONASS, Galileo, and more*. Springer Science & Business Media.

Hol, J. D. (2011). *Sensor fusion and calibration of inertial sensors, vision, ultra-wideband and GPS* (doctoral dissertation). Linköping University, Automatic Control, The Institute of Technology.

Holopainen, M., Vastaranta, M., & Hyyppä, J. (2014). Outlook for the next generation's precision forestry in Finland. *Forests*, *5*(7), 1682–1694.

Holzhüter, H., Bödewadt, J., Bayesteh, S., Aschinger, A., & Blume, H. (2023). Technical concepts of automotive lidar sensors: a review. *Optical Engineering*, *62*(3), 031213–031213.

Holzleitner, F., Langmaier, M., Hochbichler, E., Obermayer, B., Stampfer, K., & Kanzian, C. (2019). Effect of prior tree marking, thinning method and topping diameter on harvester performance in a first thinning operation-a field experiment. *Silva Fennica*, *53*(3).

Horn, B. K. P. (1986). *Robot vision*. MIT Press, Cambridge.

Hough, P. V. C. (1959). Machine analysis of bubble chamber pictures. In *Proc. of the international conference on high energy accelerators and instrumentation, sept. 1959* (pp. 554–556).

Howard, A. (2008). Real-time stereo visual odometry for autonomous ground vehicles. In *IEEE/RSJ international conference on intelligent robots and systems* (pp. 3946–3952).

Hruska, J., Čermák, J., & Šustek, S. (1999). Mapping tree root systems with ground-penetrating radar. *Tree physiology*, *19*(2), 125–130.

Hua, M.-D., Ducard, G., Hamel, T., Mahony, R., & Rudin, K. (2013). Implementation of a nonlinear attitude estimator for aerial robotic vehicles. *IEEE Transactions on Control Systems Technology*, *22*(1), 201–213.

Huck, T., Westenberger, A., Fritzsche, M., Schwarz, T., & Dietmayer, K. (2011). Precise timestamping and temporal synchronization in multi-sensor fusion. In *IEEE intelligent vehicles symposium (IV)* (pp. 242–247).

Humeau-Heurtier, A. (2019). Texture feature extraction methods: A survey. *IEEE Access*, *7*, 8975–9000.

Hyyppä, E., Hyyppä, J., Hakala, T., Kukko, A., Wulder, M. A., White, J. C., ... Kaartinen, H. (2020). Under-canopy UAV laser scanning for accurate forest field measurements. *ISPRS Journal of Photogrammetry and Remote Sensing*, *164*, 41–60.

Hyyppä, E., Kukko, A., Kaijaluoto, R., White, J. C., Wulder, M. A., Pyörälä, J., ... Hyyppä, J. (2020). Accurate derivation of stem curve and volume using backpack mobile laser scanning. *ISPRS Journal of Photogrammetry and Remote Sensing*, *161*, 246–262.

Hyyppä, J., Karjalainen, M., Liang, X., Jaakkola, A., Yu, X., Wulder, M., ... Karila, K. (2016). Remote sensing of forests from lidar and radar. In *Land resources monitoring, modeling, and mapping with remote sensing* (pp. 397–427). Taylor & Francis.

Hyyppä, J., Yu, X., Kaartinen, H., Kukko, A., Jaakkola, A., Liang, X., ...

Hyyppä, H. (2018). Forest inventory using laser scanning. In J. Shan & C. K. Toth (Eds.), *Topographic laser ranging and scanning* (pp. 379–412). Taylor & Francis.

Hyyti, H. (2009). *Simultaneous localization and forest mapping using 2D and 3D laser scanners* (Master's thesis, Helsinki University of Technology). Retrieved from http://urn.fi/URN:NBN:fi:aalto-202003102467

Hyyti, H. (2012). *Sensing through foliage* (Tech. Rep. No. 220). Metsäteho. Retrieved from http://www.metsateho.fi/wp-content/uploads/2015/02/Raportti_220_Report_Sensing_through_foliage_hh_jh.pdf

Hyyti, H. (2015). *The DCM-IMU algorithm*. https://github.com/hhyyti/dcm-imu. GitHub.

Ibraheem, N. A., Hasan, M. M., Khan, R. Z., & Mishra, P. K. (2012). Understanding color models: a review. *ARPN Journal of science and technology*, *2*(3), 265–275.

Imaging Source. (2007). *DFK 41AU02 color camera* [Fact sheet]. Retrieved from https://www.theimagingsource.com/products/industrial-cameras/usb-2.0-color/dfk41au02/

InvenSense. (2010). *ITG-3200 product specification* [Datasheet]. (PS-ITG-3200A-00-01.4, Revision: 1.4)

InvenSense. (2013). *MPU-6000 and MPU-6050 product specification* [Datasheet]. (PS-MPU-6000A-00, Revision: 3.4)

InvenSense. (2016). *MPU-9250 product specification* [Datasheet]. (PS-MPU-9250A-01, Revision: 1.1)

Jafari-Khouzani, K., & Soltanian-Zadeh, H. (2005). Rotation-invariant multiresolution texture analysis using Radon and wavelet transforms. *IEEE transactions on image processing*, *14*(6), 783–795.

Jain, R., Kasturi, R., & Schunck, B. G. (1995). *Machine vision* (Vol. 5). McGraw-Hill New York.

Jaynes, E. T. (2003). *Probability theory: The logic of science* (G. L. Bretthorst, Ed.). Cambridge University Press.

Jekeli, C. (2000). *Inertial navigation systems with geodetic applications*. Walter de Gruyter.

Johnsen, T. (2019). *Forest machine sales in Finland 2018*. Online. Retrieved from https://www.forestry.com/deep/forest-machine-sales-finland-2018

Jokelainen, V. (2010). *Recognition and localisation of young trees using machine perception* (Master's thesis, Aalto University, School of Electrical Engineering). Retrieved from http://urn.fi/URN:NBN:fi:aalto-2020122357418

Jones, J. P., & Palmer, L. A. (1987). An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. *Journal of neurophysiology*, *58*(6), 1233–1258.

Jones, R. S. (1991). *The C programmer's companion: ANSI C library functions*. Silicon Press.

Jurman, D., Jankovec, M., Kamnik, R., & Topič, M. (2007). Calibration and

data fusion solution for the miniature attitude and heading reference system. *Sensors and Actuators A: Physical*, *138*(2), 411–420.

Jutila, J., Kannas, K., & Visala, A. (2007). Tree measurement in forest by 2D laser scanning. In *International symposium on computational intelligence in robotics and automation* (pp. 491–496). doi: 10.1109/CIRA.2007.382858

Kaartinen, H., Hyyppä, J., Vastaranta, M., Kukko, A., Jaakkola, A., Yu, X., ... Hyyppä, H. (2015). Accuracy of kinematic positioning using global satellite navigation systems under forest canopies. *Forests*, *6*(9), 3218–3236. doi: 10.3390/f6093218

Kaempchen, N., & Dietmayer, K. (2003). Data synchronization strategies for multi-sensor fusion. In *Proceedings of the IEEE conference on intelligent transportation systems* (Vol. 85, pp. 1–9).

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of basic Engineering*, *82*(1), 35–45.

Kalmari, J. (2015). *Nonlinear model predictive control of a hydraulic forestry crane* (doctoral dissertation, Aalto University). Retrieved from http://urn.fi/URN:ISBN:978-952-60-6325-6

Kalmari, J., Backman, J., & Visala, A. (2014). Nonlinear model predictive control of hydraulic forestry crane with automatic sway damping. *Computers and Electronics in Agriculture*, *109*, 36–45. doi: 10.1016/j.compag.2014.09.006

Kalmari, J., Kulovesi, J., & Visala, A. (2011). Measuring log length using computer stereo vision in harvester head. *IFAC Proceedings Volumes*, *44*(1), 2226–2229.

Kalmari, J., Pihlajamäki, T., Hyyti, H., Luomaranta, M., & Visala, A. (2013). ISO 11783 compliant forest crane as a platform for automatic control. *IFAC Proceedings Volumes*, *46*(18), 164–169. doi: 10.3182/20130828-2-SF-3019.00046

Kamarainen, J.-K. (2012). Gabor features in image analysis. In *3rd international conference on image processing theory, tools and applications (IPTA)* (pp. 13–14).

Kangas, A., Astrup, R., Breidenbach, J., Fridman, J., Gobakken, T., Korhonen, K. T., ... Olsson, H. (2018). Remote sensing and forest inventories in Nordic countries–roadmap for the future. *Scandinavian journal of forest research*, *33*(4), 397–412.

Kangas, A., Haara, A., Holopainen, M., Luoma, V., Packalen, P., Packalen, T., ... Saarinen, N. (2019). Kaukokartoitukseen perustuvan metsävaratiedon hyötyanalyysi: MetKu-hankkeen loppuraportti. *Luonnonvara- ja biotalouden tutkimus*, *6*, 34.

Kankare, V., Holopainen, M., Vastaranta, M., Liang, X., Yu, X., Kaartinen, H., ... Hyyppä, J. (2017). Outlook for the single-tree-level forest inventory in Nordic countries. In I. Ivan, A. Singleton, J. Horák, & T. Inspektor (Eds.), *The rise of big spatial data* (pp. 183–195).

Springer.

Karam, S., Lehtola, V., & Vosselman, G. (2020). Strategies to integrate IMU and LiDAR SLAM for indoor mapping. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *1*, 223–230.

Kärhä, K., Hynönen, A., Laine, T., Strandström, M., Sipilä, K., Palander, T., & Rajala, P. T. (2014). Koneellinen metsänistutus ja sen tehostaminen Suomessa. *Metsätehon raportti*, *233*, 42. (in Finnish)

Kashani, A. H., Owen, W. S., Himmelman, N., Lawrence, P. D., & Hall, R. A. (2010). Laser scanner-based end-effector tracking and joint variable extraction for heavy machinery. *The International Journal of Robotics Research*, *29*(10), 1338–1352. doi: 10.1177/0278364909359316

Kashani, A. H., Owen, W. S., Lawrence, P. D., & Hall, R. A. (2007). Real-time robot joint variable extraction from a laser scanner. In *IEEE international conference on automation and logistics* (pp. 2994–2999). doi: 10.1109/ICAL.2007.4339095

Kattenborn, T., Eichel, J., & Fassnacht, F. E. (2019). Convolutional neural networks enable efficient, accurate and fine-grained segmentation of plant species and communities from high-resolution UAV imagery. *Scientific reports*, *9*(1), 1–9.

Kellner, D., Barjenbruch, M., Klappstein, J., Dickmann, J., & Dietmayer, K. (2014). Instantaneous ego-motion estimation using multiple Doppler radars. In *IEEE international conference on robotics and automation (ICRA)* (pp. 1592–1597).

Kerstens, R., Laurijssen, D., & Steckel, J. (2019). An optimized planar MIMO array approach to in-air synthetic aperture sonar. *IEEE Sensors Letters*, *3*(11), 1–4.

Khairuddin, A. R., Talib, M. S., & Haron, H. (2015). Review on simultaneous localization and mapping (SLAM). In *IEEE international conference on control system, computing and engineering (ICCSCE)* (pp. 85–90).

Khurana, A., & Nagla, K. (2020). An improved method for extrinsic calibration of tilting 2D LRF. *Journal of Intelligent & Robotic Systems*, *99*(3), 693–712.

Kim, D., & Langley, R. B. (2003). On ultrahigh-precision GPS positioning and navigation. *Navigation*, *50*(2), 103–116. doi: 10.1002/j.2161-4296.2003.tb00322.x

Kingsbury, N., & Magarey, J. (1998). Wavelet transforms in image processing. In *Signal analysis and prediction* (pp. 27–46). Springer.

Komatsu. (2019). *Ergonomics from a technical perspective, an overview of the Komatsu forest machines' technical landmarks.* Online. Retrieved from `https://www.komatsuforest.com/explore/ergonomics/from-a-technical-perspective`

Koval, L., Vaňuš, J., & Bilík, P. (2016). Distance measuring by ultrasonic

sensor. *IFAC-PapersOnLine*, *49*(25), 153–158.

Kukko, A., Andrei, C.-O., Salminen, V.-M., Kaartinen, H., Chen, Y., Rönnholm, P., ... Capek, K. (2007). Road environment mapping system of the Finnish Geodetic Institute–FGI Roamer. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *36*, 241–247.

Kukko, A., Kaartinen, H., & Hyyppä, H. (2020). Kinematic laser scanning for urban planning and infrastructure. *GIM International*, *34*(4), 31–35.

Kukko, A., Kaartinen, H., Hyyppä, J., & Chen, Y. (2012). Multiplatform mobile laser scanning: Usability and performance. *Sensors*, *12*(9), 11712–11733.

Kukko, A., Kaijaluoto, R., Kaartinen, H., Lehtola, V. V., Jaakkola, A., & Hyyppä, J. (2017). Graph SLAM correction for single scanner MLS forest data under boreal forest canopy. *ISPRS Journal of Photogrammetry and Remote Sensing*, *132*, 199–209.

Kulovesi, J. (2009). Motion vision based structure estimation in forest environment. In *IEEE/RSJ international conference on intelligent robots and systems* (pp. 1923–1928).

Kuuluvainen, T., Tahvonen, O., & Aakala, T. (2012). Even-aged and uneven-aged forest management in boreal Fennoscandia: a review. *Ambio*, *41*(7), 720–737.

Laaksonen, H. (2019). *Laserkeilausaineistoja ja ilmakuvia päivitetään jatkossa tiheämmin.* Online. Retrieved 2019-12-04, from https://www.maanmittauslaitos.fi/ajankohtaista/laserkeilausaineistoja-ja-ilmakuvia-paivitetaan-jatkossa-tiheammin

La Hera, P., & Morales, D. O. (2019). What do we observe when we equip a forestry crane with motion sensors? *Croatian Journal of Forest Engineering: Journal for Theory and Application of Forestry Engineering*, *40*(2), 259–280.

La Hera, P., Morales, D. O., & Mendoza-Trejo, O. (2021). A study case of dynamic motion primitives as a motion planning method to automate the work of forestry cranes. *Computers and Electronics in Agriculture*, *183*, 106037.

Lampinen, J., Laaksonen, J., & Oja, E. (1998). Pattern recognition. In C. T. Leondes (Ed.), *Image processing and pattern recognition* (pp. 1–59). Academic press.

Laplace, P. S. (1812). *Théorie analytique des probabilités*. Courcier.

Lawson, B. (2015). Motion sickness symptomatology and origins. In K. S. Hale & K. M. Stanney (Eds.), *Handbook of virtual environments: Design, implementation, and applications* (p. 531-600). CRC Press. doi: 10.1201/b17360

Lee, T. S. (1996). Image representation using 2D Gabor wavelets. *IEEE Transactions on pattern analysis and machine intelligence*, *18*(10),

959–971.

Li, M., & Mourikis, A. I. (2013). High-precision, consistent EKF-based visual-inertial odometry. *The International Journal of Robotics Research*, *32*(6), 690–711.

Li, Q., Yang, Y., Guo, Y., Li, W., Liu, Y., Liu, H., & Kang, Y. (2021). Performance evaluation of deep learning classification network for image features. *IEEE Access*, *9*, 9318–9333.

Li, X., Gunturk, B., & Zhang, L. (2008). Image demosaicing: A systematic survey. In *Visual communications and image processing 2008* (Vol. 6822, p. 68221J).

Li, Y., Mei, B., & Linhares-Juvenal, T. (2019). The economic contribution of the world's forest sector. *Forest Policy and Economics*, *100*, 236–253.

Liang, J., & Parks, T. W. (1996). A translation-invariant wavelet representation algorithm with applications. *IEEE transactions on signal processing*, *44*(2), 225–232.

Liang, X., Hyyppä, J., Kaartinen, H., Lehtomäki, M., Pyörälä, J., Pfeifer, N., ... Wang, Y. (2018). International benchmarking of terrestrial laser scanning approaches for forest inventories. *ISPRS Journal of Photogrammetry and Remote Sensing*, *144*, 137–179.

Liang, X., Kankare, V., Hyyppä, J., Wang, Y., Kukko, A., Haggrén, H., ... Vastaranta, M. (2016). Terrestrial laser scanning in forest inventories. *ISPRS Journal of Photogrammetry and Remote Sensing*, *115*, 63–77.

Liang, X., Kukko, A., Kaartinen, H., Hyyppä, J., Yu, X., Jaakkola, A., & Wang, Y. (2014). Possibilities of a personal laser scanning system for forest mapping and ecosystem services. *Sensors*, *14*(1), 1228–1248.

Lier, M., Korhonen, K. T., Tuomainen, T., Viitanen, J., & Mutanen, A. (2017). *Finland's forests 2017*. Online. Natural Resources Institute Finland (Luke). Retrieved from https://www.luke.fi/wp-content/uploads/2017/06/finlands-forests-facts-2017-www.pdf

Lin, L.-H., Lawrence, P. D., & Hall, R. (2010). Stereo vision based swing angle sensor for mining rope shovel. In *IEEE/RSJ international conference on intelligent robots and systems* (pp. 1714–1721). doi: 10.1109/IROS.2010.5650457

Lindroos, O., La Hera, P., & Häggström, C. (2017). Drivers of advances in mechanized timber harvesting–a selective review of technological innovation. *Croatian Journal of Forest Engineering*, *38*(2), 243–258.

Lindroos, O., Mendoza Trejo, O., La Hera, P., & Ortiz Morales, D. (2019). Advances in using robots in forestry operations. In J. Billingsley (Ed.), *Robotics and automation for improving agriculture* (pp. 233–260). Burleigh Dodds Science Publishing Limited.

Lindroos, O., Ringdahl, O., La Hera, P., Hohnloser, P., & Hellström, T. H. (2015). Estimating the position of the harvester head–a key step towards the precision forestry of the future? *Croatian Journal of Forest Engineering*, *36*(2), 147–164.

Liziniewicz, M., Ekö, P. M., & Klang, F. (2016). Effects of five tree-selection strategies when thinning spruce (Picea abies) stands: a case study in a field trail in southern Sweden. *Scandinavian journal of forest research*, *31*(5), 495–506.

Lou, L., Xu, X., Cao, J., Chen, Z., & Xu, Y. (2011). Sensor fusion-based attitude estimation using low-cost MEMS-IMU for mobile robot navigation. In *6th IEEE joint international information technology and artificial intelligence conference (ITAIC)* (Vol. 2, pp. 465–468).

Low, K.-L. (2004). Linear least-squares optimization for point-to-plane ICP surface registration. *Chapel Hill, University of North Carolina*, *4*(10), 1–3.

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, *60*(2), 91–110.

Lu, Z., Hu, Z., & Uchimura, K. (2009). SLAM estimation in dynamic outdoor environments: A review. In *International conference on intelligent robotics and applications* (pp. 255–267).

Luczak, S., Oleksiuk, W., & Bodnicki, M. (2006). Sensing tilt with MEMS accelerometers. *IEEE Sensors Journal*, *6*(6), 1669–1675.

Luinge, H. J., & Veltink, P. H. (2004). Inclination measurement of human movement using a 3-D accelerometer with autocalibration. *Transactions on neural systems and rehabilitation engineering*, *12*(1), 112–121.

Lymberopoulos, D., & Liu, J. (2017). The Microsoft indoor localization competition: Experiences and lessons learned. *IEEE Signal Processing Magazine*, *34*(5), 125–140. doi: 10.1109/MSP.2017.2713817

Lyon, R. F. (2010). Machine hearing: An emerging field. *IEEE Signal Processing Magazine*, *27*(5), 131–139.

Ma, W. J., Beck, J. M., Latham, P. E., & Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nature neuroscience*, *9*(11), 1432.

Madgwick, S. O. H., Harrison, A. J. L., & Vaidyanathan, R. (2011). Estimation of IMU and MARG orientation using a gradient descent algorithm. In *International conference on rehabilitation robotics (ICORR)* (pp. 1–7).

Magliocchetti, D., Prandi, F., Panizzoni, G., Lotto, D., & De Amicis, R. (2015). SLOPE: a 3D forest virtual system to support harvesting operations in mountain areas. In *Forest engineering: making a positive contribution. abstracts and proceedings of the 48th symposium on forest mechanization, Linz, Austria, 2015* (pp. 69–76).

Magnusson, M. (2009). *The three-dimensional normal-distributions transform: an efficient representation for registration, surface analysis, and loop detection* (doctoral dissertation). Örebro University.

Mahony, R., Hamel, T., & Pflimlin, J.-M. (2008). Nonlinear complementary filters on the special orthogonal group. *Transactions on automatic*

control, *53*(5), 1203–1218.

Majumder, S., & Deen, M. J. (2020). A robust orientation filter for wearable sensing applications. *IEEE Sensors Journal*, *20*(23), 14228–14236.

Mallat, S. (1999). *A wavelet tour of signal processing*. Elsevier.

Marjanen, K., Ojala, P., & Ihalainen, H. (2008). Measurement of annual ring width of log ends in forest machinery. In *Image processing: Algorithms and systems VI* (Vol. 6812, p. 68120D).

Marr, D., & Hildreth, E. (1980). Theory of edge detection. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, *207*(1167), 187–217.

Marshall, H. D., Murphy, G. E., & Boston, K. (2006). Evaluation of the economic impacts of length and diameter measurement error on mechanical harvesters and processors operating in pine stands. *Canadian journal of forest research*, *36*(7), 1661–1673.

Matej, J. (2014). Determination of forestry machine's tilt angle using camera and image processing. *Computers and electronics in agriculture*, *109*, 134–140.

Melkas, T., Miettinen, M., Hämäläinen, J., & Einola, K. (2014). Puukarttajärjestelmä hakkuun tehostamisessa. *Metsäteho Oy Raportti*, *230*, 20.

Metsätrans. (2017). *Metsäkonetilastot uudistuvat.* Online. Retrieved 2019-03-05, from https://metsatrans.com/metsakonetilastot-uudistuivat

Metsätrans. (2019). *2018 oli kaikkien aikojen ennätysvuosi harvesterien rekisteröinnissä.* Online. Retrieved 2019-03-05, from https://metsatrans.com/2018-oli-kaikkien-aikojen-ennatysvuosi-harvestereiden-rekisteroinnissa

Mielikäinen, J., Koskinen, I., Handroos, H., Toivanen, P., & Kälviäinen, H. (2001). Positioning of flexible boom structure using neural networks. In *Proceedings of the 9th international conference on computer analysis of images and patterns* (pp. 435–442). doi: 10.1007/3-540-44692-3_53

Miettinen, M., Kulovesi, J., Kalmari, J., & Visala, A. (2010). New measurement concept for forest harvester head. In *Field and service robotics* (pp. 35–44).

Miettinen, M., Öhman, M., Visala, A., & Forsman, P. (2007). Simultaneous localization and mapping for forest harvesters. In *Proceedings of the IEEE international conference on robotics and automation* (pp. 517–522). doi: 10.1109/ROBOT.2007.363838

Mihelj, M., Bajd, T., Ude, A., Lenarčič, J., Stanovnik, A., Munih, M., ... Šlajpah, S. (2019). *Robotics*. Springer.

Milne, B., Chen, X., Hann, C., & Parker, R. (2013). Robotisation of forestry harvesting in New Zealand–an overview. In *10th IEEE international conference on control and automation (ICCA)* (pp. 1609–1614).

Mohamed, S. A. S., Haghbayan, M.-H., Westerlund, T., Heikkonen, J., Ten-

hunen, H., & Plosila, J. (2019). A survey on odometry for autonomous navigation systems. *IEEE Access*, *7*, 97466–97486.

Mohtashami, S., Bergkvist, I., Löfgren, B., & Berg, S. (2012). A GIS approach to analyzing off-road transportation: a case study in Sweden. *Croatian Journal of Forest Engineering*, *33*(2), 275–284.

Mulligan, I. J., Mackworth, A. K., & Lawrence, P. D. (1989). A model-based vision system for manipulator position sensing. In *Workshop on interpretation of 3D scenes* (pp. 186–193). doi: 10.1109/TDSCEN.1989.68118

Muñoz-Salinas, R., & Garrido-Jurado, S. (2016). *ArUco Library*. http://sourceforge.net/projects/aruco. SourceForge.

Musso, C., Oudjane, N., & Le Gland, F. (2001). Improving regularised particle filters. In A. Doucet, N. de Freitas, & N. Gordon (Eds.), *Sequential monte carlo methods in practice* (pp. 247–271). Springer. doi: 10.1007/978-1-4757-3437-9_12

Næsset, E., Gobakken, T., Holmgren, J., Hyyppä, H., Hyyppä, J., Maltamo, M., ... Söderman, U. (2004). Laser scanning of forest resources: the Nordic experience. *Scandinavian Journal of Forest Research*, *19*(6), 482–499.

Nagatani, K., Tokunaga, N., Okada, Y., & Yoshida, K. (2008). Continuous acquisition of three-dimensional environment information for tracked vehicles on uneven terrain. In *IEEE international workshop on safety, security and rescue robotics* (pp. 25–30).

Näsi, R., Honkavaara, E., Tuominen, S., Saari, H., Pölönen, I., Hakala, T., ... Reinikainen, J. (2016). UAS based tree species identification using the novel FPI based hyperspectral cameras in visible, NIR and SWIR spectral ranges. *International archives of the photogrammetry, remote sensing and spatial information sciences; Volume XLI-B1*.

Neuvition. (2023). *Titan P1-SL905-135 solid-state high resolution lidar* [Datasheet]. (V2.1)

Nevatia, R. (1982). *Machine perception*. Prentice-Hall.

Nilsson, U., Fahlvik, N., Johansson, U., Lundström, A., & Rosvall, O. (2011). Simulation of the effect of intensive forest management on forest production in Sweden. *Forests*, *2*(1), 373–393.

NOLTFOX. (2006). *About forestry in northern European countries.* Online. Retrieved from http://noltfox.metla.fi/nordic.htm

Noordermeer, L., Sørngård, E., Astrup, R., Næsset, E., & Gobakken, T. (2021). Coupling a differential global navigation satellite system to a cut-to-length harvester operating system enables precise positioning of harvested trees. *International Journal of Forest Engineering*, *32*(2), 119–127.

Nordfjell, T., Björheden, R., Thor, M., & Wästerlund, I. (2010). Changes in technical performance, mechanical availability and prices of machines used in forest operations in Sweden from 1985 to 2010. *Scan-*

dinavian Journal of Forest Research, *25*(4), 382–389.

Nordlie, A., & Till, S. (2015). *Head-mounted displays for harvester operators–a pilot study* (Master's thesis). KTH, Industrial Engineering and Management.

Nourani-Vatani, N., & Borges, P. V. K. (2011). Correlation-based visual odometry for ground vehicles. *Journal of Field Robotics*, *28*(5), 742–768.

Novatel. (2015). *An introduction to GNSS: GPS, GLONASS, Galileo and other global navigation satellite systems* (2nd ed.). Novatel Inc.

Nvidia. (2012). *Nvidia CUDA C programming guide, version 4.2* (Tech. Rep.). Nvidia Corporation. Retrieved from https://developer.download.nvidia.com/compute/DevZone/docs/html/C/doc/CUDA_C_Programming_Guide.pdf

Oberländer, J., Pfotzer, L., Roennau, A., & Dillmann, R. (2015). Fast calibration of rotating and swivelling 3-D laser scanners exploiting measurement redundancies. In *IEEE/RSJ international conference on intelligent robots and systems (IROS)* (pp. 3038–3044).

Ocando, M. G., Certad, N., Alvarado, S., & Terrones, Á. (2017). Autonomous 2D SLAM and 3D mapping of an environment using a single 2D LIDAR and ROS. In *Latin american robotics symposium (LARS) and brazilian symposium on robotics (SBR)* (pp. 1–6).

Öhman, M., Miettinen, M., Kannas, K., Jutila, J., Visala, A., & Forsman, P. (2008). Tree measurement and simultaneous localization and mapping system for forest harvesters. In *Proceedings of the 6th international conference on field and service robotics* (pp. 369–378). doi: 10.1007/978-3-540-75404-6_35

Ohta, Y.-I., Kanade, T., & Sakai, T. (1980). Color information for region segmentation. *Computer graphics and image processing*, *13*(3), 222–241.

Ojala, T., Pietikainen, M., & Maenpaa, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on pattern analysis and machine intelligence*, *24*(7), 971–987.

Olivera, A., & Visser, R. (2016). Using the harvester on-board computer capability to move towards precision forestry. *NZ Journal of Forestry*, *60*(4), 3.

Ortiz Morales, D., Westerberg, S., La Hera, P. X., Mettin, U., Freidovich, L., & Shiriaev, A. S. (2014). Increasing the level of automation in the forestry logging process with crane trajectory planning and control. *Journal of Field Robotics*, *31*(3), 343–363. doi: 10.1002/rob.21496

Ouster. (2021). *OS1: Mid-range high-resolution imaging lidar* [Datasheet]. (REV: 2/10/2021)

Ovaskainen, H. (2009). *Timber harvester operators' working technique in first thinning and the importance of cognitive abilities on work*

productivity (doctoral dissertation). University of Helsinki.

Ovaskainen, H., Uusitalo, J., & Sassi, T. (2006). Effect of edge trees on harvester positioning in thinning. *Forest Science*, *52*(6), 659–669.

Ovaskainen, H., Uusitalo, J., & Väätäinen, K. (2004). Characteristics and significance of a harvester operators' working technique in thinnings. *International Journal of Forest Engineering*, *15*(2), 67–77.

Palander, T. S., Eronen, J. P., Peltoniemi, N. P., Aarnio, A. I., Kärhä, K., & Ovaskainen, H. K. (2019). Improving a stem-damage monitoring system for a single-grip harvester using a logistic regression model in image processing. *Biosystems engineering*, *180*, 36–49.

Palonen, T. (2016). *Augmented reality based human machine interface for semiautonomous work machines* (Master's thesis, Aalto University, School of Electrical Engineering). Retrieved from http://urn.fi/URN:NBN:fi:aalto-201611025386

Patel, M. N., & Tandel, P. (2016). A survey on feature extraction techniques for shape based object recognition. *International Journal of Computer Applications*, *137*(6), 16–20.

Patole, S. M., Torlak, M., Wang, D., & Ali, M. (2017). Automotive radars: A review of signal processing techniques. *IEEE Signal Processing Magazine*, *34*(2), 22–35.

Pedersen, H. C., Andersen, T. O., & Nielsen, B. K. (2015). Comparison of methods for modeling a hydraulic loader crane with flexible translational links. *Journal of Dynamic Systems, Measurement, and Control*, *137*(10), 101012. doi: 10.1115/1.4030801

Petovello, M. G. (2003). *Real-time integration of a tactical-grade IMU and GPS for high-accuracy positioning and navigation* (doctoral dissertation). University of Calgary, Department of Geomatics Engineering.

Pettersson, N., Fahlvik, N., Karlsson, A., & och Skogsstyrelsen, A. K. (2012). *Skogsskötselserien: Röjning*. Skogsstyrelsen.

Phillips, T. G., Green, M. E., & McAree, P. R. (2016). Is it what i think it is? is it where i think it is? using point-clouds for diagnostic testing of a digging assembly's form and pose for an autonomous mining shovel. *Journal of Field Robotics*, *33*(7), 1013–1033. doi: 10.1002/rob.21643

Phuong, N. H. Q., Kang, H.-J., Suh, Y.-S., & Ro, Y.-S. (2009). A DCM based orientation estimation algorithm with an inertial measurement unit and a magnetic compass. *Journal of Universal Computer Science*, *15*(4), 859–876.

Pierzchała, M., Giguère, P., & Astrup, R. (2018). Mapping forests using an unmanned ground vehicle with 3D LiDAR and graph-SLAM. *Computers and Electronics in Agriculture*, *145*, 217–225.

Pilli, R., Grassi, G., Kurz, W. A., Fiorese, G., & Cescatti, A. (2017). The european forest sector: past and future carbon budget and fluxes under different management scenarios. *Biogeosciences*, *14*(9), 2387–2405.

Placko, D. (2007). *Fundamentals of instrumentation and measurement*. Wiley.

Porter, R., & Canagarajah, N. (1997a). Robust rotation invariant texture classification. In *IEEE international conference on acoustics, speech, and signal processing* (Vol. 4, pp. 3157–3160).

Porter, R., & Canagarajah, N. (1997b). Robust rotation-invariant texture classification: wavelet, Gabor filter and GMRF based schemes. *IEE Proceedings-Vision, Image and Signal Processing*, *144*(3), 180–188.

Premebida, C., Ambrus, R., & Marton, Z.-C. (2018). Intelligent robotic perception systems. In E. G. Hurtado (Ed.), *Applications of mobile robots*. IntechOpen.

Premerlani, W., & Bizard, P. (2009). Direction cosine matrix IMU: Theory. *Diy Drone: USA*. Retrieved from `https://raw.githubusercontent.com/eladven/Mousquito/master/angle%20estimaion/DCMDraft2.pdf`

Pu, S., Rutzinger, M., Vosselman, G., & Elberink, S. O. (2011). Recognizing basic structures from mobile laser scanning data for road inventory studies. *ISPRS Journal of Photogrammetry and Remote Sensing*, *66*(6), S28–S39.

Purfürst, F. T., & Erler, J. (2011). The human influence on productivity in harvester operations. *International Journal of Forest Engineering*, *22*(2), 15–22.

Purfürst, T. (2022). Evaluation of static autonomous GNSS positioning accuracy using single-, dual-, and tri-frequency smartphones in forest canopy environments. *Sensors*, *22*(3), 1289.

Radon, J. (1986). On the determination of functions from their integral values along certain manifolds. *IEEE transactions on medical imaging*, *5*(4), 170–176.

Raj, T., Hanim Hashim, F., Baseri Huddin, A., Ibrahim, M. F., & Hussain, A. (2020). A survey on lidar scanning mechanisms. *Electronics*, *9*(5), 741.

Ramola, A., Shakya, A. K., & Van Pham, D. (2020). Study of statistical methods for texture analysis and their modern evolutions. *Engineering Reports*, *2*(4), e12149.

Rasinmäki, J., & Melkas, T. (2005). A method for estimating tree composition and volume using harvester data. *Scandinavian journal of forest research*, *20*(1), 85–95.

Ringdahl, O. (2007). *Techniques and algorithms for autonomous vehicles in forest environment* (licentiate thesis). Umeå University.

Ringdahl, O., Hohnloser, P., Hellström, T., Holmgren, J., & Lindroos, O. (2013). Enhanced algorithms for estimating tree trunk diameter using 2D laser scanner. *Remote Sensing*, *5*(10), 4839–4856. doi: 10.3390/rs5104839

Ringdahl, O., Lindroos, O., Hellström, T., Bergström, D., Athanassiadis, D., & Nordfjell, T. (2011). Path tracking in forest terrain by an

autonomous forwarder. *Scandinavian Journal of Forest Research*, *26*(4), 350–359.

Rish, I. (2001). An empirical study of the naive Bayes classifier. In *IJCAI workshop on empirical methods in artificial intelligence* (Vol. 3, pp. 41–46).

Rissanen, A., Guo, B., Saari, H., Näsilä, A., Mannila, R., Akujärvi, A., & Ojanen, H. (2017). VTT's fabry-perot interferometer technologies for hyperspectral imaging and mobile sensing applications. In *MOEMS and miniaturized systems XVI* (Vol. 10116, p. 101160I).

Roberts, L. G. (1963). *Machine perception of three-dimensional solids* (doctoral dissertation). Massachusetts Institute of Technology.

Rossi, G. B. (2014). *Measurement and probability: A probabilistic theory of measurement with applications*. Springer Netherlands.

Rossmann, J., Schluse, M., Schlette, C., Buecken, A., Krahwinkler, P., & Emde, M. (2009). Realization of a highly accurate mobile robot system for multi purpose precision forestry applications. In *International conference on advanced robotics* (pp. 1–6).

Rossmann, J., Schluse, M., Schlette, C., Buecken, A., Krahwinkler, P., & Emde, M. (2011). Data fusion based multi-sensor system for high accuracy forest management. *Journal of Forest Planning*, *16*(Special Issue: Multipurpose Forest Management), 263-271. doi: 10.20659/jfp.16.Special_Issue_263

Routa, J., Asikainen, A., Björheden, R., Laitila, J., & Röser, D. (2013). Forest energy procurement: state of the art in Finland and Sweden. *Wiley Interdisciplinary Reviews: Energy and Environment*, *2*(6), 602–613.

Rouveure, R., Faure, P., Marionneau, A., Rameau, P., & Moiroux-Arvis, L. (2014). A new radar sensor for cutting height measurements in tree harvesting applications. In *2nd international conference on robotics, associated high-technologies and equipment for agriculture and forestry (RHEA)* (pp. 10–p).

Rouveure, R., Monod, M. O., & Faure, P. (2008). Mapping of the environment with a high resolution ground-based radar imager. In *The 14th IEEE Mediterranean electrotechnical conference (MELECON)* (pp. 822–828).

Royo, S., & Ballesta-Garcia, M. (2019). An overview of lidar imaging systems for autonomous vehicles. *Applied Sciences*, *9*(19), 4093.

Ruetz, F., Borges, P., Suenderhauf, N., Hernández, E., & Peynot, T. (2022). Forest traversability mapping (FTM): Traversability estimation using 3D voxel-based normal distributed transform to enable forest navigation. In *IEEE/RSJ international conference on intelligent robots and systems (IROS)* (pp. 8714–8721).

Ruizenaar, M., van der Hall, E., & Weiss, M. (2013). Gyro bias estimation using a dual instrument configuration. In *Proceedings of the*

2nd CEAS specialist conference on guidance, navigation & control (EuroGNC'13).

Rusinkiewicz, S., & Levoy, M. (2001). Efficient variants of the ICP algorithm. In *Proceedings third international conference on 3-D digital imaging and modeling* (pp. 145–152).

Russell, S. J., & Norvig, P. (2016). *Artificial intelligence: A modern approach*. Pearson Education Limited.

Sahawneh, L., & Jarrah, M. A. (2008). Development and calibration of low cost MEMS IMU for UAV applications. In *5th international symposium on mechatronics and its applications (ISMA)* (pp. 1–9).

Salmivaara, A., Miettinen, M., Finér, L., Launiainen, S., Korpunen, H., Tuominen, S., ... Uusitalo, J. (2018). Wheel rut measurements by forest machine-mounted LiDAR sensors–accuracy and potential for operational applications? *International Journal of Forest Engineering*, *29*(1), 41–52.

Sandru, A., Hyyti, H., Visala, A., & Kujala, P. (2020). A complete process for shipborne sea-ice field analysis using machine vision. *IFAC-PapersOnLine*, *53*(2), 14539–14545.

Saravanan, G., Yamuna, G., & Nandhini, S. (2016). Real time implementation of RGB to HSV/HSI/HSL and its reverse color space models. In *2016 international conference on communication and signal processing (ICCSP)* (pp. 0462–0466).

Saruwatari, H., & Komura, M. (1999). Synthetic aperture sonar in air medium using a nonlinear sidelobe canceler. *Electronics and Communications in Japan (Part III: Fundamental Electronic Science)*, *82*(12), 13–25.

Scaramuzza, D., & Siegwart, R. (2008). Appearance-guided monocular omnidirectional visual odometry for outdoor ground vehicles. *IEEE transactions on robotics*, *24*(5), 1015–1026.

Schmidt, S. F. (1962). State space techniques applied to the design of a space navigation system. In *Joint automatic control conference* (pp. 1–10).

Schmidt, S. F. (1966). Application of state-space methods to navigation problems. In *Advances in control systems* (Vol. 3, pp. 293–340). Elsevier.

Schreyer, P., & Pilat, D. (2001). Measuring productivity. *OECD Economic studies*, *33*(2), 127–170.

Schulze, M. (2012). Image-based navigation of forest harvesters. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *39*, B3.

Segal, A., Haehnel, D., & Thrun, S. (2009). Generalized-ICP. In *Robotics: science and systems* (Vol. 2, p. 435).

Shan, J., & Toth, C. K. (2018). *Topographic laser ranging and scanning: Principles and processing, second edition*. Taylor & Francis.

Shan, T., Englot, B., Meyers, D., Wang, W., Ratti, C., & Rus, D. (2020). LIO-SAM: Tightly-coupled lidar inertial odometry via smoothing and mapping. In *IEEE/RSJ international conference on intelligent robots and systems (IROS)* (p. 5135-5142).

Siciliano, B., & Khatib, O. (2016). *Springer handbook of robotics*. Springer International Publishing.

SICK AG. (2006). *Laser measurement systems LMS200 to LMS291* [Technical Description]. Retrieved from `http://sicktoolbox.sourceforge.net/docs/sick-lms-technical-description.pdf`

SICK AG. (2009). *LMS100/111/120/151 laser measurement systems* [Operating Instructions]. Retrieved from `http://lars.mec.ua.pt/public/Tutorials/SickLMS151/OI_LMS100.pdf`

Sihvo, S., Virjonen, P., Nevalainen, P., & Heikkonen, J. (2018). Tree detection around forest harvester based on onboard LiDAR measurements. In *Baltic geodetic congress (BGC geomatics)* (pp. 364–367).

Silversides, C. R., & Rajala, R. (1997). *Broadaxe to flying shear*. National Museum of Science and Technology, Ottawa, Canada.

Simon, D. (2006). *Optimal state estimation: Kalman, H infinity, and nonlinear approaches*. John Wiley & Sons.

Sinha, S., Jeganathan, C., Sharma, L. K., & Nathawat, M. S. (2015). A review of radar remote sensing for biomass estimation. *International Journal of Environmental Science and Technology*, *12*(5), 1779–1792.

Smith, A. F. M., & Gelfand, A. E. (1992). Bayesian statistics without tears: a sampling–resampling perspective. *The American Statistician*, *46*(2), 84–88. doi: 10.2307/2684170

Sonka, M., Hlavac, V., & Boyle, R. (2014). *Image processing, analysis, and machine vision*. Cengage Learning.

Spero, D. J. (2004). *A review of outdoor robotics research* (Tech. Rep.). Department of Electrical and Computer Systems Engineering, Monash University.

Spero, D. J., & Jarvis, R. A. (2007). *A review of robotic SLAM* (Tech. Rep.). Department of Electrical and Computer Systems Engineering, Monash University.

Stanković, R. S., & Falkowski, B. J. (2003). The Haar wavelet transform: its status and achievements. *Computers & Electrical Engineering*, *29*(1), 25–44.

Steward, B. L., & Tian, L. F. (1999). Real-time weed detection in outdoor field conditions. In *Precision agriculture and biological quality* (Vol. 3543, pp. 266–278).

Stoyanov, T. D. (2012). *Reliable autonomous navigation in semi-structured environments using the three-dimensional normal distributions transform (3D-NDT)* (doctoral dissertation). Örebro University.

Strandström, M. (2016). Mechanized young stand management in Finland. In T. Saksa (Ed.), *Proceedings of the OSCAR workshop: Mechanized*

and efficient silviculture, November 25–26, 2015 Natural Resources Institute Finland, Suonenjoki research unit, Finland (Vol. 8, pp. 7–9).

Sucre, E. B., Tuttle, J. W., & Fox, T. R. (2011). The use of ground-penetrating radar to accurately estimate soil depth in rocky forest soils. *Forest Science*, *57*(1), 59–66.

Suger, B., Steder, B., & Burgard, W. (2015). Traversability analysis for mobile robots in outdoor environments: A semi-supervised learning approach based on 3D-lidar data. In *IEEE international conference on robotics and automation (ICRA)* (pp. 3941–3946).

Sundararajan, D. (2015). *Discrete wavelet transform: A signal processing approach*. Wiley.

Szeliski, R. (2010). *Computer vision: algorithms and applications*. Springer Science & Business Media.

Taheri, H., & Xia, Z. C. (2021). SLAM; definition and evolution. *Engineering Applications of Artificial Intelligence*, *97*, 104032.

Tang, J., Chen, Y., Kukko, A., Kaartinen, H., Jaakkola, A., Khoramshahi, E., ... Hyyppä, H. (2015). SLAM-aided stem mapping for forest inventory with small-footprint mobile LiDAR. *Forests*, *6*(12), 4588–4606.

Tang, X. (1998). Texture information in run-length matrices. *IEEE transactions on image processing*, *7*(11), 1602–1609.

Tapio. (2014). *Metsänhoidon suositukset* (O. Äijälä, A. Koistinen, J. Sved, K. Vanhatalo, & P. Väisänen, Eds.). Metsätalouden kehittämiskeskus Tapio. (in Finnish)

Tazartes, D. (2014). An historical perspective on inertial navigation systems. In *International symposium on inertial sensors and systems (ISISS)* (pp. 1–5).

Thrun, S. (1998). Learning metric-topological maps for indoor mobile robot navigation. *Artificial Intelligence*, *99*(1), 21–71.

Thrun, S. (2000). Probabilistic algorithms in robotics. *Ai Magazine*, *21*(4), 93.

Thrun, S., Burgard, W., & Fox, D. (2005). *Probabilistic robotics* (R. C. Arkin, Ed.). MIT Press.

Toiviainen, V. (2017). *Inertial measurement unit based instrumentation for pose estimation and contact detection in a forestry crane* (Master's thesis, Aalto University, School of Electrical Engineering). Retrieved from `http://urn.fi/URN:NBN:fi:aalto-201705114675`

Tomaštík, J., Chudá, J., Tunák, D., Chudỳ, F., & Kardoš, M. (2021). Advances in smartphone positioning in forests: Dual-frequency receivers and raw GNSS data. *Forestry: An International Journal of Forest Research*, *94*(2), 292–310.

Torresan, C., Berton, A., Carotenuto, F., Di Gennaro, S. F., Gioli, B., Matese, A., ... Wallace, L. (2017). Forestry applications of UAVs in Europe: a review. *International Journal of Remote Sensing*, *38*(8-10), 2427–

2447.

Toth, C., & Jóźków, G. (2016). Remote sensing platforms and sensors: A survey. *ISPRS Journal of Photogrammetry and Remote Sensing*, *115*, 22–36.

Toth, C. K., & Petrie, G. (2018). Terrestrial laser scanners. In J. Shan & C. K. Toth (Eds.), *Topographic laser ranging and scanning* (pp. 29–88). Taylor & Francis.

Uotila, K. (2017). *Optimization of early cleaning and precommercial thinning methods in juvenile stand management of Norway spruce stands* (doctoral dissertation). University of Helsinki.

Uttam, D., & Culshaw, B. (1985). Precision time domain reflectometry in optical fiber systems using a frequency modulated continuous wave ranging technique. *Journal of Lightwave Technology*, *3*(5), 971–977.

Väätäinen, K., Ylimäki, R., Lamminen, S., Siren, M., Ala-Ilomäki, J., Asikainen, A., & Anttila, P. (2011). The potential of intelligent operator assisting systems in different phases of mechanized loggings. In *Pushing the boundaries with research and innovation in forest engineering. FORMEC 2011, proceedings of the 44th international symposium on forestry mechanisation, Graz, Austria, 9-13 October 2011.*

Vanclay, J. K. (2011). Future harvest: what might forest harvesting entail 25 years hence? *Scandinavian Journal of Forest Research*, *26*(2), 183–186.

van Ginkel, M., Hendriks, C. L. L., & van Vliet, L. J. (2004). *A short introduction to the Radon and Hough transforms and how they relate to each other* (Tech. Rep.). Delft University of Technology. (Technical Report, QI-2004-01)

Vanhanen, H., Jonsson, R., Gerasimov, Y., Krankina, O., & Messieur, C. (Eds.). (2012). *Making boreal forests work for people and nature*. IUFRO Special Project on World Forest, Society and Environment (IUFRO-WFSE).

Velodyne Lidar. (2018). *VLP-16 user manual* [User manual]. (63-9243 Rev. D)

Velodyne Lidar. (2019a). *Alpha Prime: Real-time lidar sensor* [Datasheet]. (63-9679 Rev-1, VLS-128)

Velodyne Lidar. (2019b). *Puck LITE: Light weight real-time lidar sensor* [Datasheet]. (63-9286 Rev-K, VLP-16-LW)

Vestlund, K., & Hellström, T. (2006). Requirements and system design for a robot performing selective cleaning in young forest stands. *Journal of terramechanics*, *43*(4), 505–525.

Vihlman, M. (2012). *Detection, localization and species identification of young trees using machine vision* (Master's thesis, Aalto University, School of Electrical Engineering). Retrieved from `http://urn.fi/URN:NBN:fi:aalto-2020122858984`

Vihonen, J., Honkakorpi, J., Mattila, J., & Visa, A. (2013a). Geometry-aided angular acceleration sensing of rigid multi-body manipulator using mems rate gyros and linear accelerometers. In *IEEE/RSJ international conference on intelligent robots and systems* (pp. 2514–2520). doi: 10.1109/IROS.2013.6696710

Vihonen, J., Honkakorpi, J., Mattila, J., & Visa, A. (2013b). Geometry-aided MEMS motion state estimation for multi-body manipulators. In *IEEE international conference on advanced intelligent mechatronics* (pp. 341–347). doi: 10.1109/AIM.2013.6584115

Vihonen, J., Honkakorpi, J., Mattila, J., & Visa, A. (2014). Geometry-aided inversion of manipulator telescopic link length from MEMS accelerometer and rate gyro readings. In *IEEE international conference on robotics and automation* (pp. 181–186). doi: 10.1109/ICRA.2014.6906607

Vihonen, J., Honkakorpi, J., Tuominen, J., Mattila, J., & Visa, A. (2016). Linear accelerometers and rate gyros for rotary joint angle estimation of heavy-duty mobile manipulators using forward kinematic modeling. *IEEE/ASME Transactions on Mechatronics*, *21*(3), 1765–1774. doi: 10.1109/TMECH.2016.2544352

Visser, R., & Obi, O. F. (2021). Automation and robotics in forest harvesting operations: Identifying near-term opportunities. *Croatian Journal of Forest Engineering: Journal for Theory and Application of Forestry Engineering*, *42*(1), 13–24.

Vivet, D., Checchin, P., & Chapuis, R. (2013). Localization and mapping using only a rotating FMCW radar sensor. *Sensors*, *13*(4), 4527–4552.

Vosselman, G., & Maas, H.-G. (2010). *Airborne and terrestrial laser scanning*. Whittles Publishing.

Vuthea, V., & Toshiyoshi, H. (2018). A design of risley scanner for LiDAR applications. In *International conference on optical MEMS and nanophotonics (OMN)* (pp. 1–2).

Wäldchen, J., Rzanny, M., Seeland, M., & Mäder, P. (2018). Automated plant species identification–trends and future directions. *PLoS computational biology*, *14*(4), e1005993.

Waldron, K. J., & Schmiedeler, J. (2008). Kinematics. In S. Bruno & K. Oussama (Eds.), *Springer handbook of robotics* (pp. 9–33). Springer. doi: 10.1007/978-3-319-32552-1_2

Wang, D., Lia, J. H., Wang, J. L., & Li, T. (2013). A simple tree detector using laser and camera fustion. *Sensors & Transducers*, *152*(5), 137.

Wang, W. (2010). *Machine audition: Principles, algorithms and systems: Principles, algorithms and systems*. IGI Global.

Watt, P. J., & Donoghue, D. N. M. (2005). Measuring forest structure with terrestrial laser scanning. *International Journal of Remote Sensing*, *26*(7), 1437–1446.

Westerberg, S. (2014). *Semi-automating forestry machines: Motion plan-*

ning, system integration, and human-machine interaction (doctoral dissertation). Umeå University.

Westerberg, S., & Shiriaev, A. (2013). Virtual environment-based teleoperation of forestry machines: Designing future interaction methods. *Journal of Human-Robot Interaction*, *2*(3), 84–110.

Whelan, T., Johannsson, H., Kaess, M., Leonard, J. J., & McDonald, J. (2013). Robust real-time visual odometry for dense RGB-D mapping. In *IEEE international conference on robotics and automation* (pp. 5724–5731).

Wilson, A. (2021). Revision of the renewable energy directive: Fit for 55 package. *European Parliamentary Research Service*, *PE 698.781*, 1–9. Retrieved from https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698781/EPRS_BRI(2021)698781_EN.pdf

Wood, D., & Bishop, M. (2012). A novel approach to 3D laser scanning. In *Proceedings of Australasian conference on robotics and automation (ACRA)* (pp. 39–47).

Wooden, D., Malchano, M., Blankespoor, K., Howardy, A., Rizzi, A. A., & Raibert, M. (2010). Autonomous navigation for BigDog. In *IEEE international conference on robotics and automation* (pp. 4736–4741).

World Bank. (2019). *Gross domestic product (current US$).* (data retrieved from World Economy & Growth indicators, https://data.worldbank.org/indicator/NY.GDP.MKTP.CD)

Wu, Z., Sun, Z., Zhang, W., & Chen, Q. (2014). A novel approach for attitude estimation using MEMS inertial sensors. In *IEEE sensors* (pp. 1022–1025).

Wulf, O., & Wagner, B. (2003). Fast 3D scanning methods for laser measurement systems. In *International conference on control systems and computer science (CSCS14)* (pp. 2–5).

Yang, J., Vela, P., Teizer, J., & Shi, Z. (2012). Vision-based tower crane tracking for understanding construction activity. *Journal of Computing in Civil Engineering*, *28*(1), 103–112. doi: 10.1061/(ASCE)CP.1943-5487.0000242

Ye, H., Chen, Y., & Liu, M. (2019). Tightly coupled 3D lidar inertial odometry and mapping. In *International conference on robotics and automation (ICRA)* (pp. 3144–3150).

Yeong, D. J., Velasco-Hernandez, G., Barry, J., & Walsh, J. (2021). Sensor and sensor fusion technology in autonomous vehicles: A review. *Sensors*, *21*(6), 2140.

Ylikorpi, T. J., Halme, A. J., & Forsman, P. J. (2017). Dynamic modeling and obstacle-crossing capability of flexible pendulum-driven ball-shaped robots. *Robotics and Autonomous Systems*, *87*, 269–280.

Yoneda, K., Suganuma, N., Yanase, R., & Aldibaja, M. (2019). Automated driving recognition technologies for adverse weather conditions. *IATSS research*, *43*(4), 253–262.

Yoshida, T., Irie, K., Koyanagi, E., & Tomono, M. (2011). 3D laser scanner with gazing ability. In *IEEE international conference on robotics and automation* (pp. 3098–3103).

Zhang, C., Hammad, A., & Rodriguez, S. (2011). Crane pose estimation using UWB real-time location system. *Journal of Computing in Civil Engineering*, *26*(5), 625–637. doi: 10.1061/(ASCE)CP.1943-5487 .0000172

Zhang, J., & Singh, S. (2014). LOAM: Lidar odometry and mapping in real-time. In *Robotics: Science and systems* (Vol. 2, p. 9).

Zhang, J., & Singh, S. (2017). Low-drift and real-time lidar odometry and mapping. *Autonomous Robots*, *41*(2), 401–416.

Zhang, L., Yang, X., Zhang, L., Liu, X., Zhao, S., & Ma, Y. (2020). Towards automatic image exposure level assessment. *Mathematical Problems in Engineering*, *2020*, 1–14.

Zheng, Y., Liu, J., Wang, D., & Yang, R. (2012). Laser scanning measurements on trees for logging harvesting operations. *Sensors*, *12*(7), 9273–9285.

Zhou, P., Li, M., & Shen, G. (2014). Use it free: Instantly knowing your phone attitude. In *Proceedings of the 20th annual international conference on mobile computing and networking* (pp. 605–616).

Ziebinski, A., Cupek, R., Erdogan, H., & Waechter, S. (2016). A survey of ADAS technologies for the future perspective of sensor fusion. In *International conference on computational collective intelligence* (pp. 135–146).

Errata

Publication III

In Table 1, the units of axle frictions should be $\frac{\mathrm{Nm}}{\mathrm{kg\,rad/s}}$, not $\frac{1}{\mathrm{s}}$.

Publication IV

The control input vector $\mathbf{u}$ containing angular velocities was accidentally omitted from the system model in Equation (6) in Publication IV. See Section 3.6 and Equation (3.32) for correction.

Publication VI

The machine vision camera is a NET GmbH Foculus FO442C (Aegis Electronic Group, 2006), not FO422C as stated in Publication VI.

A prerequisite for increasing the autonomy of forest machinery is to provide robots with digital situational awareness, including a representation of the surrounding environment and the robot's own state in it. Therefore, this article-based dissertation proposes perception systems for autonomous or semi-autonomous forest machinery as a summary of seven publications. The work consists of several perception methods using machine vision, lidar, inertial sensors, and positioning sensors. The sensors are used together by means of probabilistic sensor fusion. The proposed methods include posture measurement of a forestry crane, positioning of a freely hanging forestry crane attachment, attitude estimation of an all-terrain vehicle, positioning a head mounted camera in a forest machine cabin, detection of young trees for point cleaning, classification of tree species, and measurement of surrounding tree stems and the ground surface underneath.

Aalto-DT 157/2023

ISBN 978-952-64-1451-5 (printed)
ISBN 978-952-64-1452-2 (pdf)
ISSN 1799-4934 (printed)
ISSN 1799-4942 (pdf)

Aalto University
School of Electrical Engineering
Department of Electrical Engineering and Automation
www.aalto.fi

BUSINESS + ECONOMY

ART + DESIGN + ARCHITECTURE

SCIENCE + TECHNOLOGY

CROSSOVER

DOCTORAL THESES