

Real-Time Processing of Range Data Focusing on Environment Reconstruction

Damien Lefloch



Real-Time Processing of Range Data Focusing on Environment Reconstruction

DISSERTATION
ZUR ERLANGUNG DES GRADES EINES DOKTORS
DER INGENIEURWISSENSCHAFTEN

VORGELEGT VON
M.Sc. DAMIEN LEFLOCH

EINGEREICHT BEI DER
NATURWISSENSCHAFTLICH-TECHNISCHEN FAKULTÄT
DER UNIVERSITÄT SIEGEN
SIEGEN 2017

BETREUER UND ERSTER GUTACHTER
PROF. DR.-ING. ANDREAS KOLB
UNIVERSITÄT SIEGEN

ZWEITER GUTACHTER
PROF. TIM WEYRICH
UNIVERSITY COLLEGE LONDON

TAG DER MÜNDLICHEN PRÜFUNG
26 JANUARY 2018

Gedruckt auf alterungsbeständigem holz- und säurefreiem Papier

Berichte aus der Informationstechnik

Damien Lefloch

**Real-Time Processing of Range Data
Focusing on Environment Reconstruction**

Shaker Verlag
Aachen 2019

Bibliographic information published by the Deutsche Nationalbibliothek

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at <http://dnb.d-nb.de>.

Zugl.: Siegen, Univ., Diss., 2018

Copyright Shaker Verlag 2019

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the publishers.

Printed in Germany.

ISBN 978-3-8440-6585-5

ISSN 1610-9406

Shaker Verlag GmbH • P.O. BOX 101818 • D-52018 Aachen

Phone: 0049/2407/9596-0 • Telefax: 0049/2407/9596-9

Internet: www.shaker.de • e-mail: info@shaker.de

Acknowledgments

This dissertation would not have been possible without the constant support of persons during all these years and this section is meant to give them proper thanking.

I am first deeply grateful to my Ph.D. supervisor Prof. Dr. Andreas Kolb, who gave me such a great opportunity to work at his research group with full financial support during all these years. Thank you for your long discussions, for your time to teach me to write high quality papers and your continuous guidance.

I am also grateful to the German Research Foundation (DFG) for funding my last years of work as part of the research training group GRK 1564 *Imaging New Modalities*.

Big thanks go to all the co-authors of my publications. In particular, to Maik Keller, the perfect colleague to work with, my curiosity on online 3-D reconstruction blossoms thanks to you. The fastest responsive person I know and always providing nice comments. You are a great person and I just have one regret which is "to never have been to one of your concert". Prof. Dr. Tim Weyrich, it was always a pleasure to collaborate and write papers with you. You always point out questions that needed to be answered in the paper before reaching its final version. I was literally amazed by your efficiency to jump on a topic and find related works so rapidly. Thomas Högg, a far-distant colleague, for his coding skills and help on designing proper Graphics Pipeline. Hamed Sarbolandi, my second office mate, thank you for the nice discussions and your help in designing some experimental setups. A special thank goes to Markus Kluge, a new colleague arriving a few months before my departure, who offers me his great support to finish my latest publication.

Furthermore, I thank my office mates, Martin Pätzold and Farhoosh Alghabi for bringing so much great conversation and funs during this long working period. Additional thanks to my Bro Martin Pätzold who guides me to the proper approach for compressing robustly unit vectors. Julian Bader, more than a colleague, who gave me a place to sleep in his flat at Siegen when I was working until very late in my office. Ulrich Schipper and Martin Lambers for their helps on teaching me how to set-up our virtual reality room. Many thanks to Hendrik Hochstetter, as Martin Pätzold and Ulrich Schipper, for the great fun playing soccer. I also thank Willi Gräfrath for his support on administrative things.

I would also thank the students I have been working with for some projects. Roberto Cespi helped to acquire data for the intensity related calibration of Time-of-Flight cameras. And Sebastian Schmitz, who writes OpenGL shaders to render blended surface elements using the elliptic weighted average approach.

Finally, deep thanks to my family for their great support. More precisely, my lovely wife Jun, you were

always supportive during a difficult time of my life and offering me many extra time slots to finish my thesis. Last but certainly not least, thanks to my father and my parents-in-law who take care of my sweet little girl Freya.

SEPT. 2017

DL.

*Real-Time Processing of Range Data
Focusing on Environment Reconstruction*

ABSTRACT

With the availability of affordable range imaging sensors, providing real-time three-dimensional information of the captured scene, new types of Computer Vision applications arise. Such applications range from designing new Human-Computer interfaces (known as Natural User Interfaces) to the generation of highly detailed reconstructions of complex scenes (for example to keep track of cultural heritage or crime scenes), to autonomous driving and augmented reality.

These depth sensors are mostly based on two efficient technologies: the structured-light principle (such as the Xbox 360 version of the Kinect camera) and the time-of-flight (ToF) principle (as cameras implemented by **pmd** technologies). When ToF cameras measure the time until the light emitted by their illumination unit is backscattered to their smart detectors, the structured-light cameras project a known light pattern onto the scene and measure the amount of distortion between the emitted light pattern and its image. Both technologies have their own advantages and weak points.

This dissertation is composed of 4 contributions. First, an efficient approach is proposed to compensate motion artifact of ToF raw images. Thereafter, a work on online three-dimensional reconstruction application has been investigated to improve the robustness of the camera tracker by segmenting moving objects. The second major contribution lies on a robust handling of noise on raw data, during the full reconstruction pipeline, proposing a new type of information fusion which considered the anisotropic nature of noise present on depth data, leading to faster convergence of high-quality reconstructions. Finally, a new method has been designed which uses surface curvature information to robustly reconstruct fine structures of small objects, as well as limiting the total error of camera drift.

ZUSAMMENFASSUNG

Durch die Verfügbarkeit von kostengünstigen Nahfeldsensoren, die 3D Daten der aufgenommenen Szene in Echtzeit erstellen, entstehen neue Anwendungen im Bereich Computer Vision. Diese Anwendungen reichen von der Erstellung neuer Mensch-Maschine-Schnittstellen (bekannt als Natural User Interfaces) über die Erstellung von sehr detaillierten Rekonstruktionen komplexer Szenen (z.B. in der Spuren an Tatorten oder Kulturstätten) bis hin zu Autonomem Fahren und Erweiterter Realität.

Diese Tiefensensoren basieren hauptsächlich auf zwei effizienten Technologien, dem: Structured-Light (SL) Prinzip (wie in der Xbox 360 Kinect Kamera) und Time-of-Flight (ToF) Prinzip (wie Kameras der Firma **pmd**technologies). Während ToF-Kameras die Zeit zwischen Lichtemission der Beleuchtungseinheit und Empfang der Rückstreuung auf dem "smart detectors" messen, projizieren SL Kameras ein bekanntes Lichtmuster in die Szene und messen die Verzerrung zwischen ausgesendetem Muster und dem resultierenden Bild. Beide Technologien haben ihre Vor- und Nachteile. Diese Dissertation besteht aus vier Beiträgen. Wir schlagen einen effizienten Ansatz vor, um Bewegungsartefakte von ToF-Rohbildern zu kompensieren. Danach arbeiten wir an 3D-Rekonstruktionsanwendungen und verbessern die Robustheit des Kameratrackings durch die Segmentierung von bewegten Objekten.

Der zweite Beitrag liegt in der robusten Handhabung von Rauschen in den Rohdaten über die ganze Verarbeitungskette der Rekonstruktion. Hier wird eine neue Art der Informationsfusion verwendet, welche die anisotropischen Eigenschaften von Rauschen in den Tiefendaten berücksichtigt und damit eine schnellere Konvergenz für hochqualitative Rekonstruktionen erzielt.

Abschließend wird eine Methode entworfen welche die Information über die Oberflächenkrümmung verwendet um auch feine Strukturen von kleinen Objekten robust zu rekonstruieren. Zusätzlich wird der Gesamtfehler des Kameradrifts eingeschränkt.

Nomenclature

CCD	Charge-Coupled Device
CMOS	Complementary Metal Oxide Semi-conductor
CUDA	Compute Unified Device Architecture
CWIM	Continuous Wave Intensity Modulation
DoF	Degree(s) of Freedom
GPU	Graphics Processing Unit
ICP	Iterative Closest Point
NIR	Near Infrared
OpenGL	Open Graphics Library
PBF	Point-Based Fusion
PCL	Point Cloud Library
RADAR	Radio Detection And Ranging
RANSAC	RAdom SAmple Consensus
RMSE	Root Mean Square Error
SDA	Standard Deviation Average
SL	Structured-Light
SLAM	Simultaneous Localization And Mapping
SNR	Signal-to-noise Ratio
SSAO	Screen-Space Ambient Occlusion
SVD	Singular Value Decomposition
ToF	Time-of-Flight
TSDF	Truncated Signed Distance Function
VGA	Video Graphics Array : display standard (640 × 480)
WLS	Weighted Least Squares

List of Symbols

px	pixel unit
x	horizontal axis of an image (unit in px)
y	vertical axis of an image (unit in px)
$\mathbf{u} = (x, y)^\top \in \mathbb{R}^2$	pixel position in the camera image
s_x, s_y	pixel size of the camera chip for both axis (unit in $\text{m}\cdot\text{px}^{-1}$)
f	camera focal (unit in m)
f_x, f_y	camera focal for both axis (unit in px)
c_x, c_y	camera principal point for both axis (unit in px)
$\mathbf{K} \in \mathbb{P}^{3 \times 2}$	intrinsic camera matrix
$t \in \mathbb{N}$	timestamp or frame ID
$\mathbf{T}^{t \rightarrow (t-1)} \in \mathbb{SE}^3$	rigid transformation between two consecutive frames
$i, j \in \mathbb{N}$	point indices
$\mathbf{p}_i^t$	i -th input point in frame t
$l \in \mathbb{N}$	iteration index or pyramid level
$\hat{\mathbf{x}} \in \mathbb{R}^n$	normalized vector ($\hat{\mathbf{x}} = \mathbf{x}/\ \mathbf{x}\ $)
$\hat{\mathbf{n}} \in \mathbb{R}^3$	normal vector
$\kappa_1, \kappa_2 \in \mathbb{R}$	first and second principal curvature values
$\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2 \in \mathbb{R}^3$	corresponding directions of curvature ($\langle \hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2 \rangle = 0$)
$\mathcal{D}^t(\mathbf{u}) \in \mathbb{R}$	input depth map
$\mathcal{W}^t(\mathbf{u}) \in \mathbb{R}$	reliability map
$\mathcal{V}^t(\mathbf{u}) \in \mathbb{R}^3$	<i>vertex map</i> of object points corresponding to $\mathbf{u}$
$\mathcal{N}_{\mathcal{V}^t(\mathbf{u})}$	local neighbourhood set $\mathcal{N}$ of point $\mathcal{V}^t(\mathbf{u})$
$ \mathcal{N} \in \mathbb{N}$	number of elements (cardinality) of set $\mathcal{N}$
$\mathcal{N}^t = \{\hat{\mathbf{n}}\}$	normal map (per-point attribute)
$\mathcal{R}^t \in \mathbb{R}$	point-radius map (per-point attribute)
$\mathcal{K}^t = \{\hat{\mathbf{e}}_1, \kappa_1, \kappa_2\}$	curvature map (per-point attribute)
$\mathcal{I}_t(\mathbf{u}) = \{i\}$	index map

I DEDICATED THIS DISSERTATION TO MY FATHER JEAN-LOUIS WHO MADE LOT OF SACRIFICES FOR OFFERING ME THE OPPORTUNITY TO PURSUE MY STUDIES, AND MY UNCLE BERNARD WHO REGRETTABLY LEFT TOO EARLY.

Contents

1	INTRODUCTION	1
1.1	Context	2
1.2	Goals and Challenges	3
1.3	Contributions	4
1.4	Thesis Outline	7
2	FUNDAMENTALS	9
2.1	Perspective Camera Model	9
2.1.1	Intrinsic Parameters	10
2.1.2	Extrinsic Parameters	13
2.2	Non-contact Range Imaging Principles	15
2.2.1	Passive methods	15
2.2.2	Active methods	16
2.3	Time-of-Flight cameras	19
2.3.1	Signal Theory	19
2.3.2	Error Sources and Characteristics	22
2.4	Depth Map Pre-processing	26
2.4.1	Range Camera Calibration	26
2.4.2	Outlier Removal	27
3	DEPTH MAP PROCESSING: MOTION ARTIFACT COMPENSATION	33
3.1	Related Works	34
3.1.1	Framerate enhancement	34
3.1.2	Detect and Repair Methods	35
3.1.3	Flow-based Correction	35

CONTENTS

3.2	The Fast Flow-based Correction Approach	36
3.3	Results	38
3.4	Discussion	43
4	INTRODUCTION TO ONLINE 3-D RECONSTRUCTION METHODS USING RANGE DATA	45
4.1	Overview	46
4.1.1	Related Works	46
4.1.2	Beyond “Basic” 3-D Reconstructions	49
4.1.3	Estimation of Surface Attributes	50
4.1.4	Camera Pose Estimation	50
4.1.5	Depth Map Fusion	52
4.1.6	The Point-Based Fusion Approach	52
4.2	Dynamic Environments	56
4.2.1	Segmentation	57
4.2.2	Dynamic-aware Model Updates	58
4.3	Discussion	61
5	CURVATURE-AWARE POINT-BASED FUSION	63
5.1	Introduction and Related Works	64
5.2	Online Surface Reconstruction	65
5.3	Depth Map Pre-processing	68
5.3.1	Point-Based Fusion Surface Attributes	68
5.3.2	Curvature Estimation	69
5.4	Camera Pose Estimation	70
5.4.1	Data Association	71
5.4.2	Minimisation	72
5.5	Local Surface Reconstruction	74
5.5.1	Quadratic Surface Patch Intersection	75
5.5.2	Blending of Quadratic Surface Intersection Points	76
5.6	Depth Map Fusion	77
5.7	Deep Index Map	79
5.7.1	Point Collisions	80
5.7.2	Screen-Space Updates	80
5.8	Results	81

CONTENTS

5.8.1	Qualitative Evaluation	83
5.8.2	Quantitative Ground-truth Evaluations	86
5.8.3	Contributors to Robustness	92
5.8.4	Scalability	93
5.8.5	Live Results	95
5.9	Discussion	95
6	ANISOTROPIC POINT-BASED FUSION	97
6.1	Introduction and Prior Work	98
6.2	Anisotropic Point-based Fusion	99
6.2.1	Anisotropy	99
6.2.2	Anisotropic Fusion	100
6.3	Implementation	101
6.4	Results	104
6.4.1	Encoding Evaluation	105
6.4.2	Anisotropic Fusion Evaluation	106
6.4.3	Performance	108
6.5	Conclusion	108
7	CONCLUSION	111
7.1	Summary	112
7.2	Outlook	113
A	APPENDIX	115
A.1	Additional details on Surface Attributes Estimation	115
A.1.1	Surface Position	115
A.1.2	Surface Normal	116
A.2	Additional details on ICP	118
A.2.1	Correspondences Search	120
A.2.2	Minimisation	122
	REFERENCES	137