

**From the Institute of Medical Informatics  
of the University of Lübeck  
Director: Prof. Dr. rer. nat. habil. Heinz Handels**

# **Geometric Deep Learning for Hand Skeleton Tracking and Gesture Recognition on 3D Point Clouds**

Dissertation  
for  
Fulfillment of Requirements for the Doctoral Degree  
of the University of Lübeck

from the Department of Computer Sciences and Technical Engineering

Submitted by  
Niklas Hermes  
from Minden

Lübeck, 2023



First referee:

Second referee:

Date of oral examination:

Approved for printing. Lübeck,



# Abstract

This is an abstract.



# Zusammenfassung

Und dies ist die Zusammenfassung.





# Zusammenfassung

Die automatisierte maschinelle Interpretation einer Handpose sowie die räumliche Modellierung der Hand halten ein vielversprechendes Potential für eine berührungslose Form der Interaktion zwischen Mensch und Maschine inne. Seit Jahrzehnten werden neue Verfahren entwickelt und neue Sensoren entworfen, um diese intuitive Form der Interaktion zu ermöglichen. Durch den bahnbrechenden Erfolg von Deep Learning Verfahren hat sich der Fokus in den letzten Jahren vor allem auf bildverarbeitende Methoden gerichtet und insbesondere Bildmodalitäten mit regelmäßiger Gitterstruktur wurden intensiv erforscht. Die Hand als anatomisch äußerst komplexes Organ manifestiert sich durch die vielen Abhängigkeiten der Muskel- und Bandapparate im Allgemeinen als anspruchsvoll zu modellierende Struktur. Insbesondere in diesem Zusammenhang stellen geometrische Domänen wie Punktwolken oder Graphen eine

Im Fokus dieser Arbeit steht die Untersuchung neuer Möglichkeiten, die sich durch die Verarbeitung geometrischer Domänen wie Punktwolken oder Graphen anhand lernender Verfahren im Bereiche der Handanalyse und Interpretation menschlicher Gestik ergeben. In diesem Zusammenhang werden verschiedene Techniken und Methoden entwickelt und vorgestellt, die im allgemeinen dem Spektrum der geometrischen Deep-Learning Verfahren entspringen und eine Vielzahl von Anwendungen im Feld der Handanalyse abdecken. Das Lernen auf Graphen wird sowohl für die Extraktion lokaler räumlicher Merkmale in Punktwolken untersucht, als auch um den globalen Kontext einer vorliegenden Struktur zu Analysieren und so die Lokalisation von Key-points zu ermöglichen. In dem Kontext gilt es ebenfalls den Informationsfluss zwischen zwei semantisch unterschiedlichen Punktwolken, durch eine geeignete Strategie zur Verknüpfung der Knoten innerhalb des Graphen, sinnvoll zu gestalten. Ein weiterer wichtiger Teil dieser Arbeit umfasst die Einbindung zeitlicher Informationen in Form einer Bewegungsanalyse in die Schätzung von Handposen. Hierzu wird eine neue Technik zur Optimierung von Verfahren zur Bestimmung des Szenenflusses für die spezifische Analyse der Bewegung des Handskeletts entwickelt. Dieser Bereich wird durch die Vorstellung eines neuen Deep-Learning Verfahrens vervollständigt, welches die Bewegungsinformationen einer Hand analysiert um die statische Schätzung einer Handpose zu präzisieren. Im abschließenden methodischen Teil der Arbeit wird untersucht inwieweit die Gestenerkennung von den extrahierten Handinformationen der vorgestellten Verfahren zur Handanalyse auf Punktwolken profitieren kann.

Durch die methodischen Anteile dieser Arbeit können die potentiellen Vorteile der Analyse dreidimensionaler geometrischer Domänen durch spezielle Deep-Learning Ver-

---

fahren herausgestellt werden. In diesem Kontext konnten besonders die mit der Hand verknüpften Problematiken von den reichen räumlichen Informationen der Modalitäten profitieren. Da ein breites Spektrum von Aufgaben innerhalb dieser Arbeit abgedeckt wurde, ist die Ergänzung vieler Bereiche durch die vorgestellten Techniken denkbar, insbesondere aber liefern die Beiträge bei Anwendungen für die gestenbasierte Mensch-Maschine Interaktion einen klaren Mehrwert.

# Contents

## References

Alldieck et al., 2017

ix



# References

- [Alldieck et al., 2017] Alldieck, T., Kassubeck, M., Wandt, B., Rosenhahn, B., and Magnor, M. “Optical Flow-Based 3D Human Motion Estimation from Monocular Video”. In: *Pattern Recognition*. Springer International Publishing, 2017, pp. 347–360.