

DISS. ETH NO. 23670

**NON-PARAMETRIC MODELS FOR
STRUCTURED DATA AND APPLICATIONS TO
HUMAN BODIES AND NATURAL SCENES**

A thesis submitted to attain the degree of
DOCTOR OF SCIENCES of ETH ZURICH
(Dr. sc. ETH Zurich)

presented by
ANDREAS M. LEHRMANN

Dipl.-Inform., University of Tuebingen

born on 16.02.1984
citizen of Germany

accepted on the recommendation of

Prof. Dr. Luc van Gool
Dr. Peter V. Gehler
Prof. Dr. Jürgen Gall

2016

Zusammenfassung

Die vorliegende Arbeit untersucht nicht-parametrische Modelle für strukturierte Daten und deren Anwendungsgebiete in der Computervision. Ziel ist die Entwicklung von kontext-sensitiven Architekturen, die zugleich expressiv und effizient sind. Unser Fokus liegt hierbei auf gerichteten graphischen Modellen, speziell Bayes'schen Netzen, in denen wir die Flexibilität nicht-parametrischer Verteilungen mit der Effizienz von Topologien mit beschränkter Baumweite kombinieren. Beschränkte Baumweiten werden entweder durch Limitierung des maximalen Eingangsgrads der unterliegenden Graphstruktur oder durch Einführung von Determinismus erzielt. Die nicht-parametrischen Verteilungen in den Knoten des Graphen sind durch Entscheidungsbäume oder Kerndichteschätzer gegeben.

Der durch spezifische Netztopologien implizierte Informationsfluss zwischen den Variablen, insbesondere die resultierenden Unabhängigkeiten, erlaubt eine natürliche Integration und Kontrolle von Kontextinformationen. Wir unterscheiden zwischen drei verschiedenen Arten von Kontext: Statisch, dynamisch und semantisch. In vier Ansätzen schlagen wir Modelle vor, die verschiedene Kombinationen dieser Kontextformen aufweisen und je nach Ausprägung die Modellierung von strukturierten Daten in Raum, Zeit und davon abgeleiteten Hierarchien erlauben. Der generative Charakter der vorgestellten Modelle ermöglicht die direkte Synthese plausibler Hypothesen.

Umfassende Experimente validieren die entwickelten Modelle in zwei Anwendungsszenarien, die in der Computervision von besonderer Bedeutung sind: Menschliche Körper und natürliche Szenen. In den praktischen Teilen dieser Arbeit behandeln wir diese beiden Großbereiche unter verschiedenen Gesichtspunkten und zeigen Anwendungen bei der Modellierung menschlicher Posen, Bewegungen und Segmentierungen sowie bei Objektkategorisierung und -lokalisierung. Wir profitieren dabei von der Verfügbarkeit moderner Datensätze mit bislang unerreichter Größe und Diversität. Der Vergleich mit klassischen Ansätzen und aktuellen Entwicklungen auf Basis etablierter Bewertungskriterien erlaubt die objektive Einordnung unserer Beiträge.

Abstract

The purpose of this thesis is the study of non-parametric models for structured data and their fields of application in computer vision. We aim at the development of context-sensitive architectures which are both expressive and efficient. Our focus is on directed graphical models, in particular Bayesian networks, where we combine the flexibility of non-parametric local distributions with the efficiency of a global topology with bounded treewidth. A bound on the treewidth is obtained by either constraining the maximum indegree of the underlying graph structure or by introducing determinism. The non-parametric distributions in the nodes of the graph are given by decision trees or kernel density estimators.

The information flow implied by specific network topologies, especially the resultant (conditional) independencies, allows for a natural integration and control of contextual information. We distinguish between three different types of context: static, dynamic, and semantic. In four different approaches we propose models which exhibit varying combinations of these contextual properties and allow modeling of structured data in space, time, and hierarchies derived thereof. The generative character of the presented models enables a direct synthesis of plausible hypotheses.

Extensive experiments validate the developed models in two application scenarios which are of particular interest in computer vision: human bodies and natural scenes. In the practical sections of this work we discuss both areas from different angles and show applications of our models to human pose, motion, and segmentation as well as object categorization and localization. Here, we benefit from the availability of modern datasets of unprecedented size and diversity. Comparisons to traditional approaches and state-of-the-art research on the basis of well-established evaluation criteria allows the objective assessment of our contributions.