Fortschritt-Berichte VDI

VDI

Reihe 10

Informatik/
Kommunikation

Dipl.-Math. Oliver Müller,
Hannover

Nr. 860

Graphical Model MAP Inference with Continuous Label Space in Computer Vision

tnt

Institut für Informationsverarbeitung
www.tnt.uni-hannover.de

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte
Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte
Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

Graphical Model MAP Inference with Continuous Label Space in Computer Vision

Der Fakultät für Elektrotechnik und Informatik

der Gottfried Wilhelm Leibniz Universität Hannover

zur Erlangung des akademischen Grades

Doktor-Ingenieur

genehmigte

Dissertation

von

Dipl. Math. Oliver Müller

geboren am 16. Juni 1986 in Berlin.

2018

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

Hauptreferent:	Prof. Dr.-Ing. Bodo Rosenhahn
Korreferent:	Prof. Dr. Bastian Leibe
Vorsitzender:	Prof. Dr.-Ing. Markus Fidler
Tag der Promotion:	20. September 2017

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

Fortschritt-Berichte VDI

Reihe 10

Informatik/
Kommunikation

Dipl.-Math. Oliver Müller,
Hannover

Nr. 860

Graphical Model MAP Inference with Continuous Label Space in Computer Vision

tnt
Institut für Informationsverarbeitung
www.tnt.uni-hannover.de

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte
Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

Müller, Oliver

Graphical Model MAP Inference with Continuous Label Space in Computer Vision

Fortschr.-Ber. VDI Reihe 10 Nr. 860. Düsseldorf: VDI Verlag 2018.
156 Seiten, 54 Bilder, 3 Tabellen.
ISBN 978-3-18-386010-4, ISSN 0178-9627,
€ 57,00/VDI-Mitgliederpreis € 51,30.

Für die Dokumentation: Maschinelles Sehen – Probabilistisch graphische Modelle – MAPInferenz – Markov-chain Monte-Carlo – Slice-Sampling – Produkt-Slice-Sampling – Artikulierte Objektverfolgung – Visuelle Objektverfolgung – Poseschätzung

This thesis deals with monocular object tracking from video sequences. The goal is to improve tracking of previously unseen non-rigid objects under severe articulations without relying on prior information such as detailed 3D models and without expensive offline training with manual annotations. The proposed framework tracks highly articulated objects by decomposing the target object into small parts and apply online tracking. Drift, which is a fundamental problem of online trackers, is reduced by incorporating image segmentation cues and by using a novel global consistency prior. Joint tracking and segmentation is formulated as a high-order probabilistic graphical model over continuous state variables. A novel inference method is proposed, called S-PBP, combining slice sampling and particle belief propagation. It is shown that slice sampling leads to fast convergence and does not rely on hyper-parameter tuning as opposed to competing approaches based on Metropolis-Hastings or heuristic samplers.

Bibliographische Information der Deutschen Bibliothek
Die Deutsche Bibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliographie; detaillierte bibliographische Daten sind im Internet unter www.dnb.de abrufbar.

Bibliographic information published by the Deutsche Bibliothek
(German National Library)
The Deutsche Bibliothek lists this publication in the Deutsche Nationalbibliographie (German National Bibliography); detailed bibliographic data is available via Internet at www.dnb.de.

© VDI Verlag GmbH · Düsseldorf 2018

Alle Rechte, auch das des auszugsweisen Nachdruckes, der auszugsweisen oder vollständigen Wiedergabe (Fotokopie, Mikrokopie), der Speicherung in Datenverarbeitungsanlagen, im Internet und das der Übersetzung, vorbehalten.

Als Manuskript gedruckt. Printed in Germany.

ISSN 0178-9627

ISBN 978-3-18-386010-4

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

Acknowledgement

This thesis was written in the course of my activity as a scientific research assistant at the Institut für Informationsverarbeitung (TNT), Leibniz Universität Hannover.

First of all, I would like to thank my doctoral advisor Prof. Dr.-Ing. Bodo Rosenhahn for giving me the opportunity to work in his group, for many inspiring discussions, and for his continuing support and guidance. Many thanks are due to Prof. Dr. Bastian Leibe for being my second reviewer, and Prof. Dr.-Ing. Markus Fidler for serving as chair of my thesis defense committee.

I would also thank Prof. Dr.-Ing. Jörn Ostermann and Prof. Dr.-Ing. Bodo Rosenhahn for providing an outstanding working environment at the TNT.

Special thanks go to all of my colleagues at the TNT, who make this lab such a fun place to work. I especially would like to thank Dr.-Ing. Kai Cordes for his guidance during my study and diploma thesis, which motivated me to start working towards a doctoral degree. I also thank Prof. Dr.-Ing. Michael Ying Yang for his support and for sharing his deep knowledge on probabilistic graphical models with me. I further thank my officemate Karsten Vogt for many inspiring, and constructive discussions.

Special thanks to Matthias Schuh, Dr.-Ing. Martin Pahl, Dr.-Ing. Marco Munderloh, and Thomas Wehberg for their outstanding technical and administrative support, as well as to the secretaries, Hilke Brodersen, Doris Jaspers-Göring, Melanie Huch, and Pia Bank for their invaluable administrative work. This institute would not have been such an outstanding place to work without their continuing efforts.

Last but not least, special thanks go to my family and my friends. Without the unconditional support of my parents Rosel and Michael Müller, this work would not have been possible.

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte
Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte
Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

Contents

1 Introduction **1**
1.1 Motivation . . . 2
1.2 Contributions . . . 5
1.3 Structure of the thesis . . . 7
1.4 List of Publications . . . 10

2 Related Work **15**
2.1 Object Tracking . . . 16
2.2 Approximate MAP Inference in Probabilistic Graphical Models . . . 19

3 Fundamentals **25**
3.1 Notations . . . 26
3.2 Probabilistic Graphical Models . . . 27
3.2.1 Bayes Network . . . 28
3.2.2 Markov Random Field . . . 28
3.2.3 Factor Graph . . . 31
3.3 Inference in Probabilistic Graphical Models . . . 33
3.3.1 Max-Product Belief Propagation . . . 35
3.3.2 Max-Product Loopy Belief Propagation . . . 37
3.3.3 Dual Methods . . . 39
3.4 Markov Chain Monte-Carlo Methods . . . 45
3.4.1 Metropolis-Hastings Sampler . . . 48
3.4.2 Gibbs Sampler . . . 48
3.4.3 Slice Sampler . . . 50
3.5 Max-Product Particle Belief Propagation . . . 54

4 Stochastic Inference in Probabilistic Graphical Models **56**
4.1 Slice-Sampling Particle Max-Product . . . 59
4.1.1 Sampling from Product-Max Distributions . . . 59
4.1.2 Particle Max-Product . . . 61
4.1.3 Computing the Slice Regions . . . 64
4.1.4 Random Walk Analysis . . . 66
4.2 Tree-Reweighted Particle Max-Product . . . 69
4.2.1 Diverse Particle Selection . . . 71
4.2.2 Heuristic Proposals versus Slice-Sampling . . . 73

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte
Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

4.3 Discussion 76

5 Tracking 78
5.1 Part-Based Template Tracking 80
5.1.1 Inference 81
5.1.2 Experiments 82
5.1.3 Discussion 84
5.2 Object Tracking Demonstrator 85
5.3 Joint Tracking and Segmentation 88
5.3.1 Joint Pose Estimation and Segmentation 91
5.3.2 Dual Decomposition 93
5.3.3 Visibility Estimation 96
5.3.4 Experiments 99
5.3.5 Summary 103

6 Conclusions 107
6.1 Future Work 109

A Appendix 111
A.1 Optical Coherence Tomography 112
A.1.1 Problem Setup 112
A.1.2 Motion Compensation 113
A.1.3 OCT Image Undistortion 121

Bibliography 128

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte
Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

Abbreviations

2D	two-dimensional
3D	three-dimensional
ADMM	alternating direction method of multipliers
BP	belief propagation
DAG	directed acyclic graph
DD	dual decomposition
DPM	deformable parts model
DPMP	diverse particle max-product
FMP	flexible mixtures-of-parts
HOG	histogram of oriented gradients
KL	Kullback-Leibler divergence
LP	linear program
MAP	maximum a posteriori
MATLAB®	Matrix Laboratory, a proprietary programming language and IDE developed by MathWorks
MCMC	Markov chain Monte-Carlo
MH	Metropolis-Hastings
MH-PBP	Metropolis-Hastings particle belief propagation
MP-BP	max-product belief propagation
MRF	Markov random field
MuPAD®	a computer algebra system bundled with MATLAB®
OCT	optical coherence tomography
OTB	online tracking benchmark
PBP	particle belief propagation
PCP	percentage of correct parts
PGM	probabilistic graphical model
RGB	red, green, and blue

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

RMSD	root-mean-square deviation
S-PBP	slice-sampling particle belief propagation
SVM	support vector machine
TRBP	tree-reweighted belief propagation
VDPM	visibility-aware deformable parts model
VDPM-e	visibility-aware deformable parts model without edge
VOT	visual object tracking

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte
Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

Symbols and Notation

Symbols for Probability Theory

$\mathcal{N}(x;\mu,\sigma)$	normal distribution with mean μ and standard deviation σ
Ω	sample space
P	probability distribution
$p(x)$	probability density function
$\mathbf{X}$	random variable vectors
X, Y, Z	random variables
$\mathcal{X},\mathcal{Y},\mathcal{Z}$	state space
x	random variable values
$\mathbf{x}$, $\mathbf{y}$, $\mathbf{z}$	random variable value vectors

Symbols for Probabilistic Graphical Models

$\mathcal{C}$	set of cliques
c	clique
$\mathbf{d}$	data
$\mathcal{E}$	edges
$E(x)$	energy function
$\mathcal{F}$	factor vertices
$\mathcal{G}$	graphical model
$g(\lambda)$	dual function
k, l	state indices

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte
Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

$\mathcal{L}$	local polytope
L_s	number of states for discrete state space $\mathcal{X}_s$
$\mathcal{L}(\{\mathbf{y}_\tau\}_\tau,\mathbf{y},\lambda)$	(augmented) Lagrangian function
$\lambda \in \Lambda$	Lagrange multipliers
$\mathcal{M}$	marginal polytope
$m_{t\to s}(x_s)$	belief propagation message from vertex t to s
$\overline{m}_{t\to s}(x_t)$	pre-message from vertex t to s
$\hat{m}_{t\to s}(x_s)$	approximate belief propagation message from vertex t to s
$\hat{\mu}_s(x_s)$	approximate max-marginal function over variable x_s
$\mu_s(x_s)$	max-marginal function over variable x_s
$\mathcal{N}$	neighborhood system
$n = 1, ..., N$	BP/DD/ADMM iteration
$\mathrm{Pa}(s)$	parents of vertex s
$\mathcal{P}_s$	particle set for vertex s
$\phi_s(x_s), \phi_{s,t}(x_s, x_t), \phi_c(x_c)$	unary potential, pairwise potential, and clique potential
$\psi_s(x_s), \psi_{s,t}(x_s, x_t), \psi_c(x_c)$	unary energy, pairwise energy, and clique energy
r, s, t	vertex indices
$\mathcal{S}$	visiting schedule
$\tau \in \mathcal{T}$	subproblems
θ	parameter vector
$\mathcal{V}$	random variable vertices
Z	partition function

Symbols for MCMC

A	slice

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

$i = 1, ..., p$	particle index
$m = 1, ..., M$	Markov chain index
$\mu(x)$	MCMC target density function
$\nu_0(x)$	MCMC initial density function
$\nu(x)$	marginal probability
$x_s^{(i)\langle m\rangle}$	particle at the m-th MCMC iteration
$\bar{x}_s^{(i)\langle m\rangle}$	candidate particle
$q(x \mid y)$	MCMC proposal density function
$q_\sigma(x \mid y)$	Gaussian proposal density with mean y and standard deviation σ
$T(x, y)$	MCMC transition kernel
$\mathcal{U}$	auxiliary state space
$\mathcal{U}(A)$	uniform distribution over region A
u	auxiliary variable
Z	normalization constant
$\{u^{\langle m\rangle}\}_m$	Markov chain auxiliary variables
$\{x_s^{(i)}\}_i$	particle
$\{x^{\langle m\rangle}\}_m$	Markov chain variables

Symbols for Part-based Object Tracking

$E_{\mathrm{con}}(\mathbf{p}, \mathbf{y})$	global consistency energy
$E_{\mathrm{DPM}}(\mathbf{p})$	deformable parts model energy
$E_{\mathrm{seg}}(\mathbf{y})$	segmentation energy
$E_{\mathrm{VDPM}}(\mathbf{p}, \mathbf{v})$	visibility-aware deformable parts model energy
$\mathbb{I}[\,\cdot\,]$	indicator function
$\mu(\mathbf{p})$	patch mask function

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

$\boldsymbol{\mu}$	patch mask matrix
$\mathbf{p}$	vector of object poses
$\mathbf{v}$	visibility v_i for each DPM patch i
$\mathbf{x}$	indicator vector over object poses
$\mathcal{X}^{\text{pose}}$	pose state space
$\mathcal{X}^{\text{seg}}$	segmentation state space
$\mathbf{y} \in \{0,1\}^N$	foreground/background segmentation of an image with N pixels

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte
Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

Abstract

This thesis deals with the development and improvement of maximum a posteriori (MAP) inference approaches in probabilistic graphical models (PGMs) and their application on challenging computer vision problems.

Many challenging computer vision tasks are modeled as MAP inference problems in PGMs. MAP inference is the problem of finding the most probable configuration of random variables for a given target problem in the exponentially large space of possible outcomes. PGMs are a family of powerful modeling languages which unify two fundamental concepts: uncertainty and graphical models. Many real-world phenomena can be modeled in form of probability distributions over continuous-valued random variables. A PGM is a language to model these distributions which typically involve a very large number of random variables. Conditional independence assumptions of the random variables play a key role in retrieving tractable models.

In the first part of this thesis, a general purpose framework for MAP inference in PGMs over continuous-valued random variables based on stochastic inference methods is developed. A novel approach, the slice-sampling particle belief propagation (S-PBP) algorithm, is developed which achieves more accurate and faster MAP estimates than heuristic sampling or Metropolis-Hastings sampling approaches. The proposed approach generates sample proposals from the max-marginal distributions using the slice sampling algorithm. By exploiting the message-passing nature of the applied MAP inference approach, the dependence on hyper-parameters is reduced and a significant speedup is achieved.

The second part of this thesis is dedicated to the application of the developed inference approaches to computer vision applications. Hereby, the main focus is in online tracking of articulated objects. The visual tracking of previously unseen objects in videos or video streams is a fundamental task in computer vision. A novel framework is proposed for part-based object tracking. The problem of automatic model initialization and the reduction of tracker drift by incorporating higher-order constraints and image segmentation cues to the tracker is addressed. A global consistency prior is proposed which enables inference of both part-based tracking and image segmentation in a joint probabilistic model. Experiments show that the joint formulation leads to improved image segmentation results as well as reduced drift in online object tracking.

Keywords: computer vision, probabilistic graphical models, MAP inference, Markov-chain Monte-Carlo, slice sampling, product slice sampling, articulated online tracking, visual object tracking, pose estimation

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

Kurzfassung

Das Ziel dieser Arbeit ist die Erstellung und Verbesserung von Optimierungsverfahren zur Inferenz in probabilistischen graphischen Modellen (PGMs) und deren Anwendung auf Probleme im Bereich Computer-Vision.

Viele Computer-Vision Probleme werden heutzutage als MAP-Inferenz Probleme in PGMs behandelt. MAP-Inferenz beschreibt hierbei das Finden der wahrscheinlichsten Kombination von Werten im exponentiell wachsenden Raum möglicher Lösungen eines Problems. Probabilistische graphische Modelle sind eine Familie von Modellierungssprachen zur Beschreibung von Verbundwahrscheinlichkei-ten über einer Menge von Zufallsvariablen. Hierbei werden zwei grundlegende Prinzipien miteinander vereint: Die Modellierung von Unsicherheiten und die Modellierung (bedingter) Unabhängigkeit von Zufallsvariablen mittels Knoten und Kanten in einem Graph.

Der erste Teil dieser Arbeit behandelt das Problem der MAP-Inferenz bei reellwertigen Zufallsvariablen mit Hilfe stochastischer Inferenzverfahren. Slice-sampling particle belief propagation (S-PBP) ist ein neu entwickelter Ansatz, welcher eine genauere MAP-Schätzung in kürzerer Zeit erlaubt als andere stochastische Verfahren. Eine Kernkomponente stochastischer Suchverfahren ist die Erzeugung von Stichproben im Lösungsraum. Bisherige Verfahren sind entweder heuristisch motiviert oder verwenden Vorschlagsverteilungen deren Parameter Problem-abhängig eingestellt werden müssen. Der in dieser Arbeit vorgestellte Ansatz erzeugt Stichproben direkt aus den Max-Marginal-Verteilungen des graphischen Modells mit Hilfe des Slice-Sampling Verfahrens. Durch Ausnutzung des Message-Passing Mechanismus der verwendeten Optimierungsverfahren wird die Parameter-Abhängigkeit verringert und eine Beschleunigung des Verfahrens erreicht.

Im zweiten Teil der Arbeit werden die zuvor entwickelten Inferenzverfahren auf Probleme im Bereich Computer-Vision angewandt. Das Hauptaugenmerk liegt hierbei auf der Online-Verfolgung artikulierter Objekte in Videosequenzen. Die visuelle Verfolgung zuvor unbekannter Objekte in Videos ist ein fundamentales Problem des maschinellen Sehens. Es wird ein neuer Ansatz zur Teile-basierten Objektverfolgung entwickelt. Das Problem der automatischen Modellinitialisierung und der Reduktion des Driftens in Teile-basierten Modellen wird durch die Integration von Bildsegmentierung und zusätzlicher Bedingungen höherer Ordnung behandelt. Es wird ein globaler Konsistenzterm vorgeschlagen, der die Teile-basierte Poseschätzung und die Bildsegmentierung in einem gemeinsamen probabilistischen Modell vereint. Experimente zeigen das die gemeinsame Schätzung von Pose und Segmentierung zum einen die Bildsegmentierung verbessert, als auch das Driften der geschätzten Pose effektiv verringert.

Stichworte: maschinelles Sehen, Probabilistische graphische Modelle, MAP-Inferenz, Markov-chain Monte-Carlo, Slice-Sampling, Produkt-Slice-Sampling, Artikulierte Objektverfolgung, Visuelle Objektverfolgung, Poseschätzung

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

Chapter 1

Introduction

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte
Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

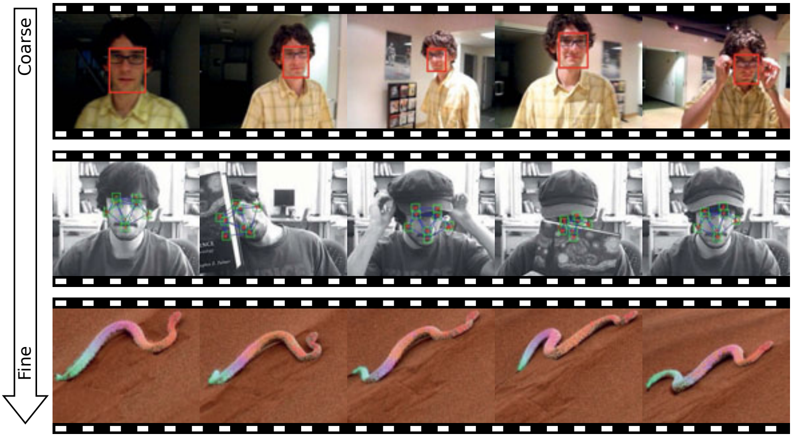


Figure 1.1: Object tracking with varying level of detail. Top: Tracking of the coarse extents of an object (the bounding box, visualized as a red rectangle). Middle: Tracking of an object which is decomposed into smaller parts (green rectangles) which are related to each other (blue lines). Bottom: Fine-scale part-based tracking of articulated objects (parts are color-coded from the head to tail fading from red over blue to green).

1.1 Motivation

The field of computer vision deals with constructing theories and algorithms towards automatic processing and interpretation of visual data. The amount of generated video data has steadily grown over the last decades due to the availability of cheap sensors and the presence of fast networks. For example, three hundred hours of video was uploaded per minute on the popular online video platform YouTube in 2016 [20]. Comparing to four years ago, the upload rate increased by 400% [21].

One of the most fundamental problems in computer vision is the problem of following a single object in a video, also known as visual object tracking. As shown in Fig. 1.1, object tracking can be handled in different granularities. From coarse single bounding-box estimates, over semantic parts decomposition, up to fine-grained pixel-wise tracking.

Using special sensor systems such as the Kinect [103] which fuse visual cameras with depth sensors, it is already possible to robustly extract human pose estimates with precise positions of the limbs in real time [43]. These systems have a limited sensor range and do not work outdoors with direct sunlight due to its infra-red pattern projection technique. Furthermore, these systems require a large database

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

of training images with manually annotated poses. The construction of a training database is expensive and already existing datasets may not contain the target object types or have incompatible groundtruth annotations (such as bounding-box annotations instead of part annotations).

The problem of tracking arbitrary objects with fine-detailed position estimates of all parts without relying on restrictive offline training is widely unsolved. Prior methods are either detector-based and hence require as large as possible training database with manually annotated part poses [35], are restricted to coarse bounding-box estimates only [70], or use registration-based methods to track the object surface [100]. The latter approach is limited to (in-)extensible surfaces. The handling of occlusions or the expression of articulated motion is a major problem in registration-based approaches.

In this thesis, part-based approaches are developed for tracking of articulated objects under the regime of PGMs. The focus here is on online-tracking based approaches. That is, the object appearance is unknown a priori and has to be learnt from a manual initialization in the first video frame.

The contribution of this thesis is two-fold. First, efficient algorithms are presented to perform inference in PGMs. PGMs are a family of powerful languages for modeling complex systems over a large number of variables using the notion of probabilistic uncertainty and (conditional) independence assumptions. Hereby, the focus in this thesis is on retrieving the most likely configuration of unobserved (or hidden) variables depending on the observed variables (the input data). This is also known as MAP inference. The second contribution of this thesis is to apply the developed methods for inference in PGMs to computer vision problems.

The object tracking problem is an instance of the class of *inverse problems*. Given an observation $\mathbf{d} = (d_1, \ldots, d_M)$ (for instance, a video sequence) and a model of the underlying problem f (the object tracking model) one wants to infer the model parameter values $\mathbf{x} = (x_1, \ldots, x_N)$ (the object part poses) which describe the observation most likely:

$$\mathbf{x} = f^{-1}(\mathbf{d}). \tag{1.1}$$

This is in contrast to *forward problems* which map model parameters to (possibly perturbed) observations:

$$\mathbf{d} = f(\mathbf{x}) + \nu \tag{1.2}$$

with noise ν. Many inverse problems of interest (especially in computer vision) are *ill-posed* [42]. That is, a solution $\mathbf{x}$ of Eq. (1.1) might not be unique and may jump with slightly perturbed input $\mathbf{d}$. This is an undesired property since, for example, one might not expect a completely different pose estimate when only changing a single pixel in the image. Ill-posed problems can be (approximately) transformed into well-posed problems by introducing the concept of *regularization* [110].

A related issue is that observations $\mathbf{d}$ are often noisy and corrupted (for example, thermal noise of the image sensors and compression artifacts). Furthermore due to

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

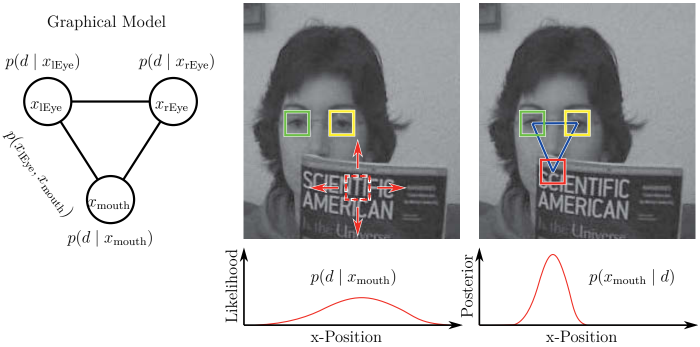

Graphical Model
$p(d \mid x_{\text{lEye}})$
$p(d \mid x_{\text{rEye}})$
x_{lEye}
x_{rEye}
$p(x_{\text{lEye}}, x_{\text{mouth}})$
x_{mouth}
$p(d \mid x_{\text{mouth}})$
Likelihood
$p(d \mid x_{\text{mouth}})$
x-Position
Posterior
$p(x_{\text{mouth}} \mid d)$
x-Position

Figure 1.2: Left: probabilistic graphical model of a part-based face localization application. Middle: the part likelihood (bottom) of the mouth location (red box). Right: the posterior (bottom) of the mouth location (red box) under consideration of all part locations (green and yellow box) and prior information about the relative part locations (blue lines).

computational considerations, the model under consideration can only be a simplistic approximation of the underlying true creation process. All these factors induce *uncertainty* to the inference process. Uncertainty can be modeled via the introduction of *probabilities*. The probabilistic counterpart to Eq. (1.1) is the maximum likelihood estimator:

$$\max_{x_1,\ldots,x_N} p(d_1,\ldots,d_M \mid x_1,\ldots,x_N). \tag{1.3}$$

Regularization then corresponds to introducing a prior $p(x_1,\ldots,x_N)$, leading to the MAP estimator:

$$\max_{x_1,\ldots,x_N} p(d_1,\ldots,d_M \mid x_1,\ldots,x_N)p(x_1,\ldots,x_N). \tag{1.4}$$

In PGMs [61], complex global relationships between the latent variables are modeled via simpler local variable interactions. Graph notations such as nodes and edges are used to model *independence assumptions* which correspond to a factorization of the joint distribution into a set of local functions $p(\mathbf{x} \mid \mathbf{d}) = \prod_c \phi_c(\mathbf{x}_c \mid \mathbf{d})$. Figure 1.2 shows an example graphical model for part-based face localization in a single frame. The posterior is proportional to the product of a *likelihood* term $p(\mathbf{d} \mid \mathbf{x})$ consisting of independent part-wise appearance terms and a *prior* $p(\mathbf{x})$ consisting of pairwise functions encoding the spatial relationships of the parts.

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

The part likelihoods can have a complex, multi-modal shape, for instance due to miss-detections, occlusions, or image noise. A parametrization via uni-modal Gaussian distributions is hence inappropriate. Furthermore, a naive discretization of the search space suffers from low-quality approximation of the original continuous objective function with respect to the number of discretization steps. A large number of discretization steps are required to cover the regions of high probability. On the other hand, many steps are wasted in the (often much larger) regions of low probability.

In this thesis, stochastic inference methods based on Markov chain Monte-Carlo (MCMC) simulation [3, 88] are applied for MAP inference in PGMs with continuous state space. For MCMC, the generation of sample proposals is of fundamental importance for efficient state space exploration. We show that the structure of the MAP inference method can be exploited for generating high-informative sample proposals and hence to accelerate MAP inference in continuous state space.

1.2 Contributions

The main contributions of this thesis are grouped in two parts:

- Stochastic inference in graphical models with continuous variables, and
- online part-based object tracking.

In the first part, the particle max-product belief propagation framework is studied and methods for faster state space exploration are developed. An alternative, more efficient particle sampling method based on product slice-sampling is proposed. In the second part, the proposed inference methods are applied to the task of part-based object tracking. A framework for object tracking based on probabilistic graphical models is developed and extended by (semi-) automatic model initialization and occlusion reasoning towards long-term object tracking. The two contributions are summarized in the following.

Stochastic Inference in Graphical Models with Continuous Variables. Message passing mechanisms are at the heart of many variational inference approaches in probabilistic graphical models. Hereby, messages are iteratively sent between nodes and edges of a graph until an equilibrium is reached. Since the messages themself are functions and can be of arbitrary shape, a compact (approximate) representation is required. In this work, particle sampling is used to approximate the messages. The approximation quality highly depends on the particle sampling procedure. Usually, heuristic sampling or Metropolis-Hastings based Markov chain Monte Carlo simulation is used to generate particle proposals. Hereby, new particles are sampled from a conditional distribution, which is *not* the target distribution.

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

This sampling process needs to be repeated several times in order to retrieve samples from the desired target distribution. The number of required iterations is called *burn-in period* and highly depends on the *correlation* of the sampled particles. Reducing this correlation is at the heart of effective particle sampling.

We propose to use an alternative approach for particle sampling, called slice sampling. This method is more robust towards hyper-parameter selection than Metropolis-Hastings sampling. Additionally to that, we show how to exploit the structure of the message passing scheme and apply product slice sampling instead of black-box sampling. This eliminates the dependence on hyper-parameters completely. The resulting approach, referred to as slice-sampling particle belief propagation (S-PBP), leads to faster convergence with a shorter burn-in period. We further show increased empirical performance on an image denoising problem.

As a second contribution, a diverse particle selection method is integrated in the S-PBP framework. We show that diverse particle selection with S-PBP proposals leads to higher-probable MAP estimates in less iterations than other heuristic sample proposal methods.

Summary of contributions:

- The slice-sampling particle belief propagation (S-PBP) approach is developed and compared to Metropolis-Hastings sampling.
- S-PBP is combined with a diverse particle selection approach. This method performs favorably towards heuristic sample proposals.

Part-Based Object Tracking. The following challenges in online multi-part object tracking are addressed: (i) Tracking of previously unseen objects without resorting to costly offline-training, (ii) automatic model initialization from a single reference image, and (iii) occlusion reasoning for each object part.

The first part is addressed by using a deformable parts model (DPM) with HOG features and template matching. By using the S-PBP approach for MAP inference, we obtain highly accurate tracking results.

Based on this approach, a real-time tracking system is developed and integrated in a demonstrator application. Hereby, a simple game is controlled only by visual input from a consumer webcam.

The second part deals with automatic model initialization. Hereby, a foreground/background mask is used to automatically construct a (multi-scale) graphical model.

Occlusion reasoning is handled by proposing a novel global visibility prior. The standard deformable parts model is extended by auxiliary variables encoding the visibility of each part. Tree-structured deformable parts models suffer from the well-known double-counting problem. Due to the decoupling of the model parts, it often happens that certain image areas are actually explained twice by two parts, whereby other image areas are completely ignored. The global visibility prior can

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

prevent such double-counting by enforcing that an image area can be fully covered by at most one part. Furthermore, the global visibility prior allows the incorporation of foreground/background segmentation cues.

The proposed method is applied on challenging self-recorded sequences to perform articulated object tracking of highly deformable objects. Furthermore, we evaluate our method on a state-of-the-art online object tracking benchmark showing comparable performance to state-of-the-art discriminative trackers.

Summary of contributions:

- Fine-grained, multi-part object tracking of previously unseen objects.
- Automatic model initialization from foreground mask.
- Joint pose estimation, image segmentation, and visibility reasoning using a global shape prior.

1.3 Structure of the thesis

In the following, the structure of this thesis is summarized. A graphical overview is given in Fig. 1.3.

Chapter 2 The related work in (multi-part) object tracking is summarized. In a second part, the state-of-the-art of inference in probabilistic graphical models is shortly reviewed and the various concepts of exploring large state spaces are discussed.

Chapter 3 Relevant basics for this thesis is introduced in this chapter. This includes a condensed introduction to probabilistic graphical models and relevant techniques for marginal and MAP inference. Furthermore, an introduction to particle sampling strategies in graphical models is given.

Chapter 4 This chapter contains the first part of main contributions of this thesis. A novel inference algorithm called *slice-sampling particle belief propagation (S-PBP)* is introduced. It is shown that the structure of max-product messages in the belief propagation framework enables efficient particle sampling using the *product slice sampler* which is introduced in the previous chapter. The performance of the proposed methods is evaluated empirically on synthetic datasets. The MCMC random walk behavior of S-PBP and MH-PBP proposals is compared. Furthermore, S-PBP and heuristic proposals are compared in a *diverse particle selection* framework.

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

Chapter 5 This chapter introduces a *multi-part object tracking* framework. The inference techniques of the previous chapter are applied. First, it is shown how graphical models can be used for *object tracking with manual initialization.* The second contribution of this chapter is an approach based on a foreground/background segmentation mask which allows *semi-automatic initialization* of the object tracking framework. We propose a *global visibility prior* for joint image segmentation, tracking, and part visibility estimation. Graph decomposition methods are combined with S-PBP which allows efficient inference in the high-order graphical models. Our proposed multi-part object tracker framework is evaluated on a state-of-the-art online tracking benchmark dataset as well as on self-recorded, challenging video sequences of highly articulated objects.

Chapter 6 Contributions and discussions of the previous chapters are summarized and concluded. Interesting directions for future research are given.

Appendix A A use case for medical image processing is presented in the appendix on the topic of motion compensation and image undistortion of 3D optical coherence tomography (OCT) images. The motion compensation and image undistortion problems are formulated as MAP inference problems in continuous space PGMs. Motion compensation is solved via approximation as a Gaussian MRF. Image undistortion is obtained by skin surface segmentation (using a MRF) and fitting of an a priori known percutaneous implant model.

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

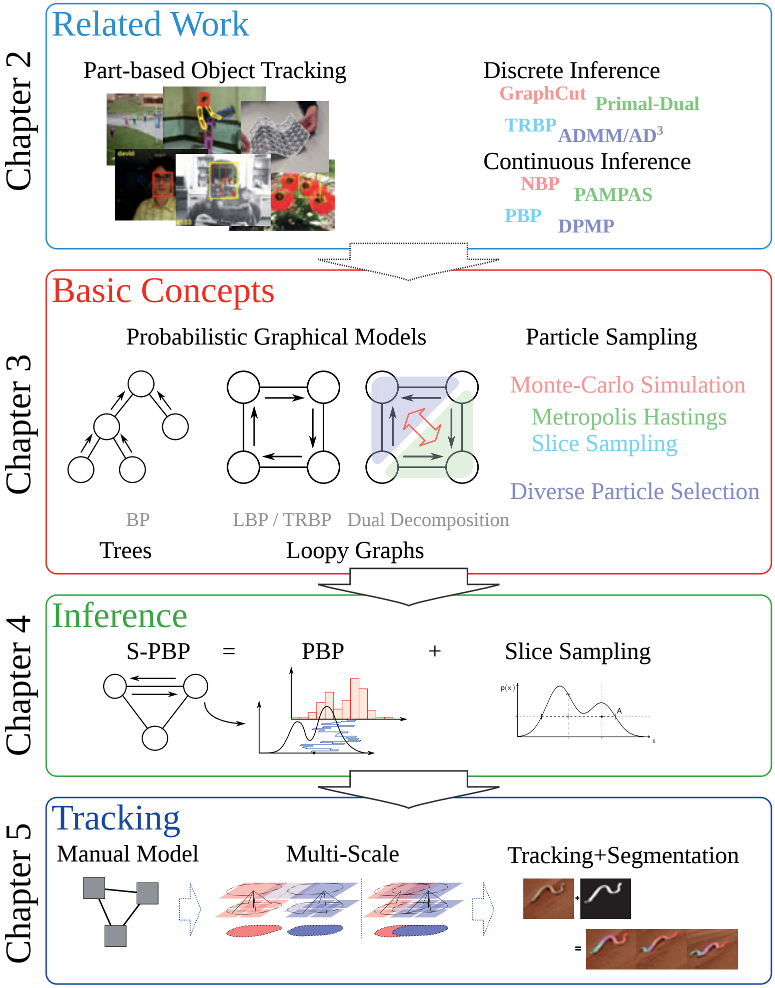


Figure 1.3: Thesis overview.

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

1.4 List of Publications

The following papers have been published by the author during the time of the dissertation. The first two papers cover the topic of non-rigid online object tracking. The focus is set on developing efficient inference algorithms in continuous state PGMs. Chapters 4 and 5 is based on these publications. The last three publications are in cooperation with the Laser Zentrum Hannover (LZH) and handle the topic of image undistortion and motion compensation of in-vivo OCT scans. These papers concentrate on the modeling aspect rather than the development of sophisticated inference techniques. This work is summarized in Appendix A.1.

[87] **Oliver Müller**, Michael Y. Yang, Bodo Rosenhahn. Slice Sampling Particle Belief Propagation. *In: Proc. of the IEEE International Conference on Computer Vision (ICCV)*, December 2013.

Inference in continuous label Markov random fields is a challenging task. We use particle belief propagation (PBP) for solving the inference problem in continuous label space. Sampling particles from the belief distribution is typically done by using Metropolis-Hastings (MH) Markov chain Monte Carlo (MCMC) methods which involves sampling from a proposal distribution. This proposal distribution has to be carefully designed depending on the particular model and input data to achieve fast convergence. We propose to avoid dependence on a proposal distribution by introducing a slice sampling based PBP algorithm. The proposed approach shows superior convergence performance on an image denoising toy example. Our findings are validated on a challenging relational 2D feature tracking application.

[86] **Oliver Müller**, Bodo Rosenhahn. Global Consistency Priors for Joint Part-based Object Tracking and Image Segmentation (Accepted). *Winter Conference on Applications of Computer Vision (WACV)*, 2017.

Tracking of previously unseen, articulated objects is an active research area. Recently, deformable parts model (DPM) have been used to improve the online tracking performance for bounding-box trackers. We extend the DPM with global priors which enforce consistency with foreground/background segmentation cues. We propose a Dual Decomposition approach and show how to efficiently solve the high-order coupling constraints as a feasible sub-problem. The proposed approach is evaluated on the VOT online tracking benchmark, outperforming the baseline in both tracking accuracy and robustness. We further show that in presence of stable image segmentation cues, the flexibility of a generic DPM generated from a single reference frame can be improved by introducing the concept of *part visibility*, the visibility-aware DPM (VDPM). This allows for fine-grained articulated object tracking using an automatically generated DPM from a single template image.

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

[84] **Oliver Müller**, Sabine Donner, Tobias Klinder, Ralf Dragon, Ivonne Bartsch, Frank Witte, Alexander Krüger, Alexander Heisterkamp, Bodo Rosenhahn. Model Based 3D Segmentation and OCT Image Undistortion of Percutaneous Implants. *In: Proc. of Medical Image Computing and Computer-Assisted Intervention, 14th International Conference (MICCAI), Lecture Notes in Computer Science (LNCS)*, September 2011.

Optical Coherence Tomography (OCT) is a noninvasive imaging technique which is used here for in vivo biocompatibility studies of percutaneous implants. A prerequisite for a morphometric analysis of the OCT images is the correction of optical distortions caused by the index of refraction in the tissue. We propose a fully automatic approach for 3D segmentation of percutaneous implants using Markov random fields with application to refractive image undistortion. Refraction correction is done by using the subcutaneous implant base as a prior for model based estimation of the refractive index using a generalized Hough transform. Experiments show the competitiveness of our algorithm towards manual segmentations done by experts.

[85] **Oliver Müller**, Sabine Donner, Tobias Klinder, Ivonne Bartsch, Alexander Krüger, Alexander Heisterkamp, Bodo Rosenhahn. Compensating motion artifacts of 3D in vivo SD-OCT scans. *In: Proc. of Medical Image Computing and Computer-Assisted Intervention, 15th International Conference (MICCAI), Lecture Notes in Computer Science (LNCS)*, October 2012.

We propose a probabilistic approach for compensating motion artifacts in 3D in vivo SD-OCT (spectral-domain optical coherence tomography) tomographs. Subject movement causing axial image shifting is a major problem for in vivo imaging. Our technique is applied to analyze the tissue at percutaneous implants recorded with SD-OCT in 3D. The key challenge is to distinguish between motion and the natural 3D spatial structure of the scanned subject. To achieve this, the motion estimation problem is formulated as a conditional random field (CRF). For efficient inference, the CRF is approximated by a Gaussian Markov random field. The method is verified on synthetic datasets and applied on noisy in vivo recordings showing significant reduction of motion artifacts while preserving the tissue geometry.

[29] Sabine Donner, **Oliver Müller**, Frank Witte, Ivonne Bartsch, Elmar Willbold, Tammo Ripken, Alexander Heisterkamp, Bodo Rosenhahn, Alexander Krüger. In situ optical coherence tomography of percutaneous implant-tissue interfaces in a murine model. *Biomedical Engineering, De Gruyter*, Karlsruhe, May 2013.

Novel surface coatings of percutaneous implants need to be tested in biocompatibility studies. The use of animal models for testing usually involves

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

numerous lethal biopsies for the analysis of the implant-tissue interface. In this study, optical coherence tomography (OCT) was used to monitor the reaction of the skin to a percutaneous implant in an animal model of hairless but immunocompetent mice. In vivo optical biopsies with OCT were taken at days 7 and 21 after implantation and post mortem on the day of noticeable inflammation. A Fourier-domain OCT was programmed for spoke pattern scanning schemes centered at the implant midpoint to reduce motion artifacts during in vivo imaging. Image segmentation allowed the automatic detection and morphometric analysis of the skin contour and the subcutaneous implant anchor. On the basis of the segmentation, the overall refractive index of the tissue within one OCT data set was estimated as a free parameter of a fitting algorithm, which corrects for the curved distortion of the planar implant base in the OCT images. OCT in combination with the spoke scanning scheme and image processing provided time-resolved three-dimensional optical biopsies around the implants to assess tissue morphology.

Additional papers from collaborations with group members.

[114] Karsten Vogt, **Oliver Müller**, Jörn Ostermann. Facial Landmark Localization using Robust Relationship Priors and Approximative Gibbs Sampling. *In: Proc. of Advances in Visual Computing*, December 2015.

We tackle the facial landmark localization problem as an inference problem over a Markov Random Field. Efficient inference is implemented using Gibbs sampling with approximated full conditional distributions in a latent variable model. This approximation allows us to improve the runtime performance 1000-fold over classical formulations with no perceptible loss in accuracy. The exceptional robustness of our method is realized by utilizing a L1 -loss function and via our new robust shape model based on pairwise topological constraints. Compared with competing methods, our algorithm does not require any prior knowledge or initial guess about the location, scale or pose of the face.

[23] Kai Cordes, **Oliver Müller**, Bodo Rosenhahn, Jörn Ostermann. HALF-SIFT: High-Accurate Localized Features for SIFT. *In: Proc. of the 22nd IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Workshop on Feature Detectors and Descriptors: The State Of The Art and Beyond*, June 2009.

In this paper, the accuracy of feature points in images detected by the Scale Invariant Featuer Transform (SIFT) is analyzed. It is shown that there is a systematic error in the feature point localization. The systematic error is caused by the improper subpel and subscale estimation, an interpolation with a parabolic function. To avoid the systematic error, the detection of High-Accurate Localized Features (HALF) is proposed. We present two models

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

for the localization of a feature point in the scale-space, a Gaussian and a Difference of Gaussians based model function. For evaluation, ground truth image data is synthesized to experimentally prove the systematic error of SIFT and to show that the error is eliminated using HALF. Experiments with natural image data show that the proposed methods increase the accuracy of the feature point positions by 13.9 % using the Gaussian and by 15.6 % using the Difference of Gaussians model.

[24] Kai Cordes, **Oliver Müller**, Bodo Rosenhahn, Jörn Ostermann. Bivariate Feature Localization for SIFT Assuming a Gaussian Feature Shape. *In: Proc. of Advances in Visual Computing, 7th International Symposium (ISVC), Lecture Notes in Computer Science (LNCS)*, November 2010.

In this paper, the well-known SIFT detector is extended with a bivariate feature localization. This is done by using function models that assume a Gaussian feature shape for the detected features. As function models we propose (a) a bivariate Gaussian and (b) a Difference of Gaussians. The proposed detector has all properties of SIFT, but provides invariance to affine transformations and blurring. It shows superior performance for strong viewpoint changes compared to the original SIFT. Compared to the most accurate affine invariant detectors, it provides competitive results for the standard test scenarios while performing superior in case of motion blur in video sequences.

[25] Kai Cordes, **Oliver Müller**, Bodo Rosenhahn, Jörn Ostermann. Feature Trajectory Retrieval with Application to Accurate Structure and Motion Recovery. *In: Proc. of Advances in Visual Computing, 7th International Symposium (ISVC), Lecture Notes in Computer Science (LNCS)*, September 2011.

Common techniques in structure from motion do not explicitly handle foreground occlusions and disocclusions, leading to several trajectories of a single 3D point. Hence, different discontinued trajectories induce a set of (more inaccurate) 3D points instead of a single 3D point, so that it is highly desirable to enforce long continuous trajectories which automatically bridge occlusions after a re-identification step. The solution proposed in this paper is to connect features in the current image to trajectories which discontinued earlier during the tracking. This is done using a correspondence analysis which is designed for wide baselines and an outlier elimination strategy using the epipolar geometry. The reference to the 3D object points can be used as a new constraint in the bundle adjustment. The feature localization is done using the SIFT detector extended by a Gaussian approximation of the gradient image signal. This technique provides the robustness of SIFT coupled with increased localization accuracy.

Our results show that the reconstruction can be drastically improved and the drift is reduced, especially in sequences with occlusions resulting from

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

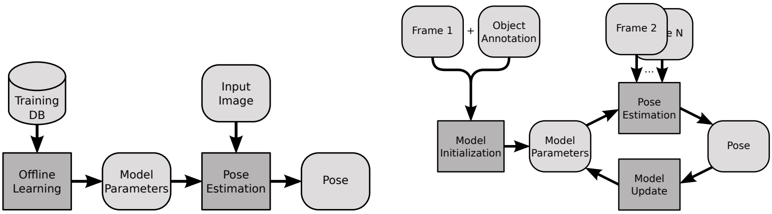
foreground objects. In scenarios with large occlusions, the new approach leads to reliable and accurate results while a standard reference method fails.

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte
Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

Chapter 2

Related Work

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte
Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.



(a) (b)

Figure 2.1: Workflow of (a) discriminative pose estimation methods and (b) online tracking methods.

The overview of related work starts with the state-of-the-art in part-based object tracking and pose estimation in section 2.1. Following that, the related work in discrete and continuous MAP inference in probabilistic graphical models is summarized in section 2.2. These approaches are finally set into context with the proposed methods in this thesis.

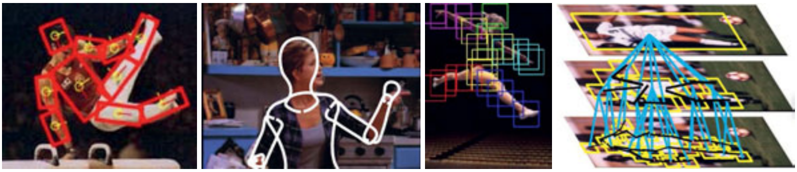
2.1 Object Tracking

Visual object tracking is a fundamental problem in computer vision. The goal is to estimate the state of a target in each frame of a given video or image sequence. Hereby, the target can be either manually selected in the first frame of the sequence [124, 69], or can be automatically detected using previously trained object detectors.

Object tracking and pose estimation are highly related computer vision tasks. Pose estimation is the problem of retrieving the location and orientation (i.e., the pose) of each body part of a target object from a single image. Object trackers work on image sequences or videos, where the same (previously annotated or automatically detected) target is to be followed. We start by examining relevant literature in the field of single frame pose estimation. Following this, the literature in object tracking is summarized and grouped in three categories: Multi-object tracking, online tracking, and articulated tracking.

Deformable part-based models for pose estimation. The literature in deformable parts models (DPMs) is vast [37, 95, 104, 60, 120, 4, 39, 119, 125, 31, 34, 102, 74, 93, 126, 135, 94, 32] and an extensive discussion of the progress this topic has made over the last decades is out of scope. A survey of part-based pose estimation methods specialized to human poses can be found in [78]. The majority of approaches are based on pictorial structures [37]. Here, a target object is decom-

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.



Pictorial Structures [4]
Deformable Structures [135, 134]
Mixtures of Parts [126]
Multi-Layer MoP [31]

Figure 2.2: Pose estimation models of varying granularity.

posed in semantically meaningful object parts which are connected by spring-like forces [37, 104, 120, 4, 39, 119, 34, 93, 78] (cf. Fig. 2.2, the leftmost picture). The flexible mixtures of parts model [125, 31, 102, 74, 126, 32] relaxes the semantic correspondence to body parts by approximating a body limb by a mixture model of smaller part patches (Fig. 2.2, the two rightmost pictures).

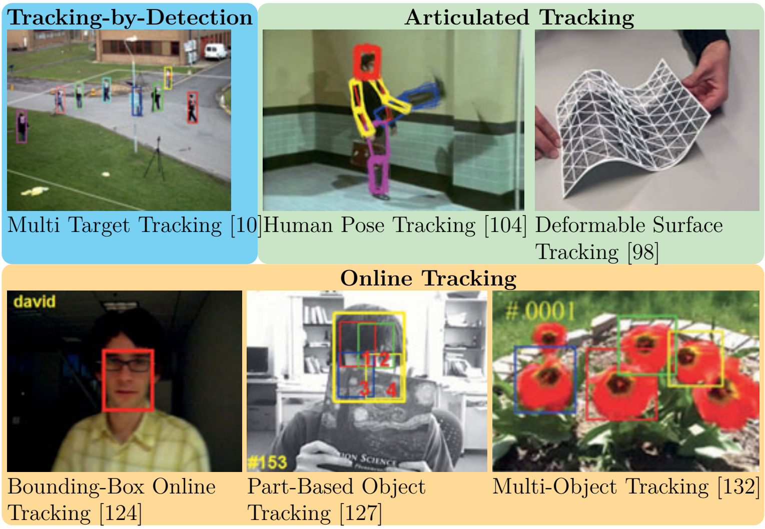
DPMs are driven by local appearance models which provide a score specifying how likely a particular part (for instance a body limb) is in a certain pose in the target image (likelihood). The part pose can be parametrized for instance by an (x,y) center position, orientation, scale, foreshortening and so on. Furthermore, contextual information in form of a prior is used to constrain the relative part positions to each other. The priors often include kinematic constraints (the left lower arm should be connected to the left upper arm). The various implementations vary in the model parts semantics, the part appearance features, as well as the part pose priors.

DPMs usually require a carefully designed supervised training on a huge set of manually annotated images (cf. the pose estimation workflow in Fig. 2.1a). This limits their application to pre-trained object classes only.

The following three paragraphs summarize relevant literature in visual object tracking. An overview is given in Fig. 2.3.

Multi-object tracking. Multi-object tracking is the process of tracking multiple instances of an object class (for instance, tracking humans on a street which is under video surveillance). Multi-object tracking is usually tackled using the tracking-by-detection paradigm, whereby the object tracking problem is decomposed into two separate steps: detection and association [10]. The detection step is applied on each video frame independently. The result is a list of object *candidates*. The association step associates each detection candidate to a single individual, leading to temporally consistent object trajectories over all video frames. The object pose is encoded as a bounding box. The use of part-based detectors, as introduced in the previous

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.



Multi Target Tracking [10]

Human Pose Tracking [104]

Deformable Surface Tracking [98]

Bounding-Box Online Tracking [124]

Part-Based Object Tracking [127]

Multi-Object Tracking [132]

Figure 2.3: Object tracking models of varying granularity and with different parts semantics.

paragraph, can significantly improve object tracking [108].

Online tracking. Online tracking is the discipline of tracking arbitrary objects agnostic of its class while learning the object appearance *during* tracking. A tracker is usually initialized on the first frame by manually providing a bounding box annotation of the target object. Online trackers are based on the causality principle. The input video sequence is processed only in a single pass (cf. Fig. 2.1b). Pose estimates are based on previous frames only without considering future frames. Unlike online tracking, batch-processing methods (as introduced in the previous paragraph) assume that the video data is already completely recorded and all time steps are available per random access. Online trackers process data from streaming devices such as consumer webcams. Hence, they are often designed with real-time constraints in mind.

Most online tracking frameworks up to date are bounding-box based, i.e. the object location is encoded as a (possibly rotated and scaled) bounding box [5, 44, 56, 129, 133, 127, 48, 132, 11]. Consequently, state-of-the-art benchmarks are restricted to bounding-box annotations [124, 71, 70]. Going beyond bounding-boxes, the recent

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

works of [127, 132, 30] are part-based approaches. They combine part-based models (see paragraph above) with online tracking. The number of parts is quite limited and the objective of interest is still in predicting bounding-boxes (cf. Fig. 2.3, bottom row, middle).

Figure 2.4 shows sequences and corresponding bounding-box annotations of the visual object tracking (VOT) 2016 benchmark from Kristan et al. [70]. The benchmark consists of 60 manually annotated video sequences. The benchmark is open source and written in MATLAB® and C. In contrast to the online tracking benchmark (OTB) benchmark [124], the sequences in VOT are selected from a broad set of sources with a focus on diversity of visual attributes such as occlusions, camera motion, illumination changes and many more. The VOT metrics handle loss of track in a more robust way than OTB by re-initializing the tracker when the estimated bounding-box has zero overlap with the groundtruth. The VOT benchmark is updated on a yearly basis since 2013 in conjunction with the VOT challenge [69].

Articulated tracking. For complex-shaped objects, bounding-box estimates are not adequate. One often is interested in estimating finer-grained articulations of these objects. Typical examples are human pose tracking [104] and hand pose tracking [107]. Another instance is tracking of (smooth) deformable surfaces [99, 98]. All these approaches have in common that they are limited to a single object class. In Chapt. 5 we propose approaches which are *agnostic* of the object class. We use part-based models with a a large set of parts. Inference in those models is challenging and requires sophisticated optimization techniques. This is the topic of the following section.

2.2 Approximate MAP Inference in Probabilistic Graphical Models

The models described in the previous section can be encoded as probabilistic graphical models (PGMs) with a continuous state space. Retrieving the most probable pose or object trajectory can then be formulated as maximum a posteriori (MAP) estimation in the PGM.

Exact inference in continuous PGMs is a rather rare case. One such case is when the posterior distribution is a (multi-variate) Gaussian and can be represented as a Gaussian Markov random field [97]. Here, tractable exact inference can be achieved using the Gaussian belief propagation algorithm [15]. In some cases, such a strict model assumption can work quite well, as shown in a medical imaging application dealing with motion compensation of 3D OCT scans as summarized in Appendix A.

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

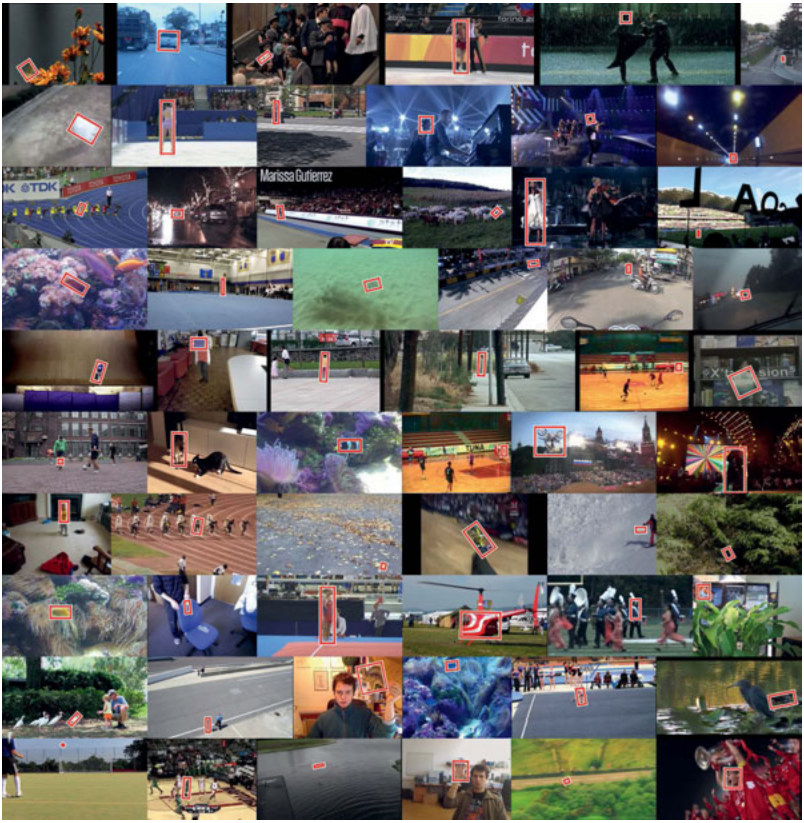


Figure 2.4: Video sequences and corresponding groundtruth annotations of the bounding box based visual object tracking benchmark in [70].

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

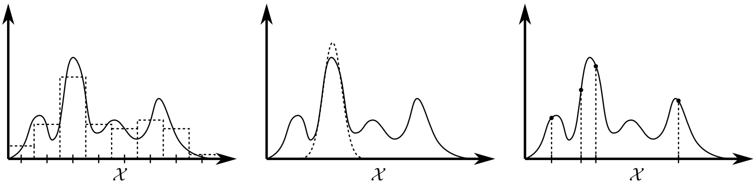


Figure 2.5: Illustration of approximation approaches, where the solid curve is the true objective function and dashed lines represent the respective function approximations. From left to right: state space discretization, variational methods, and particle sampling.

Message-passing methods. A huge class for discrete, as well as continuous inference methods in PGMs are based on *message passing.* Messages are iteratively sent between neighboring nodes in order to propagate information throughout the whole graph. This can be illustrated as every node having its own preference of being in a certain state (the marginal belief). The nodes propagate their preference by sending out messages to neighboring nodes. Simultaneously, they receive messages from its neighboring nodes and update their belief. If the update does not change the belief anymore, an equilibrium is reached. One can show that in certain cases, this belief then corresponds to the (max-)marginal distribution.

This approach has gained high popularity in the machine learning community since the *global* inference task is split into *local* operations. This has several advantages. It enables, e.g., distributed reasoning [107], parallelization and distributed computing [101] of large scale data.

Convergent message-passing approaches have strong connections to linear program relaxations and dual decomposition methods (cf. Fig. 4 in [57], Fig. 2.6). This gives them a strong theoretical foundation.

Approximate inference approaches for continuous state models can be roughly divided into three types as illustrated in Fig. 2.5: State space discretization and pruning, variational inference, and particle sampling. These three groups are shortly reviewed in the following.

Discretization and pruning. Many computer vision problems are formulated as discrete inference problems, although the underlying quantities of interest may live in a continuous space. Some typical examples are: human pose estimation and object tracking (searching for the limb's or object's position, orientation, and scale), image reconstruction (where the pixel intensities are reconstructed from noisy or blurred pixel observations).

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

Reasons for choosing discrete inference algorithms in favor of continuous methods are that the former have become computationally highly efficient. There exist certain optimality guarantees (at least for suitable relaxations of the original problem). They can handle a huge amount (in the order of 10^3–10^5) of discrete variables [57]. Often, effective heuristics can be applied for label pruning, significantly reducing the search space. E.g. the pictorial structures framework [37, 4] exploits some assumptions of the model such as limitation to tree-structured graphs, Gaussian pairwise potentials, and grid-structured search space discretization. Straightforward extension to the 3D search space is either on the border of tractability due to huge memory consumption [22], or uses sampling to reduce the search space [2, 8]. The higher-order graph matching method from [131] uses pre-filtering of matching candidates for state space pruning.

Fig. 2.6 shows an overview of recent discrete MAP inference methods in probabilistic graphical models.

Variational methods. Approaches working directly in the continuous domain are often better suited for exploiting the state space while limiting the computational load. There are two main branches: variational approximation methods, and particle sampling methods. Variational approaches (see, e.g. [117]) aim at locally approximating the objective function with a function from a much easier to handle function space. Often, the Kullback-Leibler (KL) divergence is used as a similarity measure for quantifying the similarity of the target function to its approximation. Instances of this approach are the variational message passing method of [123] and the expectation propagation framework [82].

While being computationally efficient with a low memory footprint, expectation propagation and variational methods often oversimplify the target objective function and perform local optimization. Hence, they tend to get trapped at suboptimal local modes.

A recent approach guides the variational message-passing method of [123] using pre-trained random regression forests [54]. This approach overcomes the problem of getting trapped in poor local optima and accelerates convergence, although at the expense of requiring computationally expensive training.

Particle sampling. Particle sampling based inference methods evaluate the target objective function at a finite number of points. These approaches are better suited for distributions with multiple strong modes.

The inference quality crucially depends on the generated particles in terms of *diversity* and *accuracy*. Diversity refers to the intuition that particles should ideally be spread over the whole state space in order to capture all areas with high probability. Furthermore, the particles should be positioned as close as possible to the modes of the objective function. This is referred to as accuracy in this thesis.

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

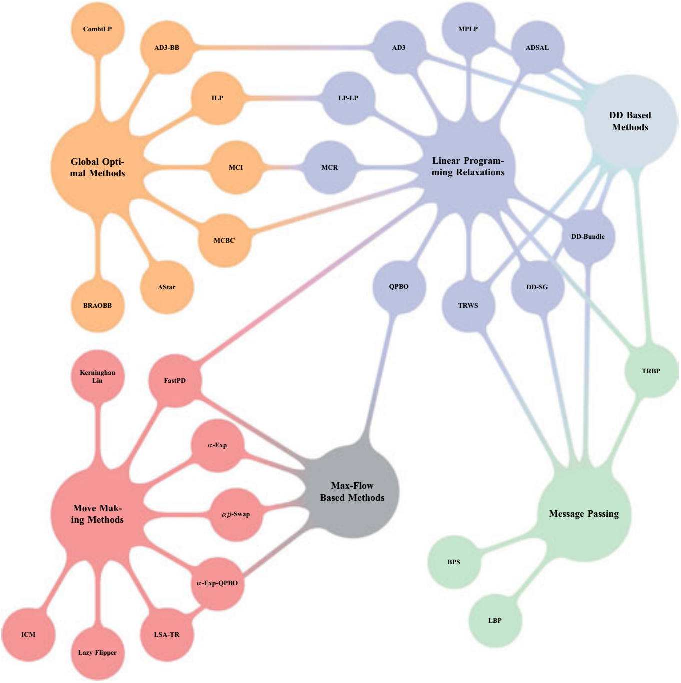


Figure 2.6: Grouping of various discrete inference algorithms (from Kappes et al. [57], Fig. 4).

In the literature, there exist many approaches for generating particle proposals.

Early methods perform particle sampling over the joint state space [28]. This has the drawback that a huge amount of particles is typically needed to capture all modes in the very high-dimensional space. Recent approaches perform particle sampling on a *local*, rather than global scale. An independent set of particles is generated for each random variable. Since the state space for each random variable is much lower than the joint state space, a significantly smaller number of particles per random variable is required.

Recent sampling approaches in graphical models exploit message passing. Here, particle sampling is used in order to approximate the messages sent between edges and nodes [62, 53, 50, 106, 67, 92, 90]. Sampling particles from belief estimates

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

using Monte-Carlo simulation was firstly proposed by Koller et al. [62]. Later on, Isard et al. [53] and Sudderth et al. [105, 107, 106] independently of each other developed approaches which approximate the messages by Gaussian mixtures and use Monte-Carlo integration to perform approximate sum-product message passing. The drawback of these approaches is that they can produce inconsistent estimates. That is, the estimated belief does not asymptotically converge to the true belief with increased number of particles. In contrast, the particle belief propagation algorithm [50, 111] produces consistent estimates. Hereby, particles are directly sampled from the (estimated) belief using Metropolis-Hastings sampling. Later on, the particle belief propagation method was adapted to max-product message passing, which is better suited for producing MAP estimates [67]. A greedy-like approach for producing MAP estimates was proposed by [112]. Here, new particles are generated by adding random noise on the current MAP estimate. A similar procedure is used by [92]. The expectation particle belief propagation method from [77] combines expectation propagation with particle belief propagation. Hereby, samples are generated from Gaussian proposal distributions as in [50]. The difference is that the proposal is dynamically adapted using expectation propagation such as to approximate the current belief estimate. A recently developed approach builds on the max-product message passing method and aims for generating a *diverse* set of particles in order to improve MAP estimates [91, 90]. They iteratively enrich the current set of particles by new (as much as possible diversely sampled) particle candidates and afterwards filter out redundant particles by considering the approximation error of the messages generated by the remaining particle set with respect to the full particle set. Their approach is highly effective in producing diverse particles but they rely on high-quality sample proposals which are generated from a combination of different heuristics. These heuristics are application dependent and require parameter tuning.

We build upon the max-product belief propagation (MP-BP) framework of [67] in Sect. 4.1 and propose a novel MCMC sampling method based on product slice sampling, called S-PBP. Our approach is more robust against hyper-parameter selection than competing methods. Furthermore, in Sect. 4.2.1 we extend the diverse sampling framework of [90] with S-PBP which replaces the heuristic proposal generators. We show that S-PBP leads to lower energy estimates while achieving faster convergence.

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

Chapter 3

Fundamentals

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte
Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

This chapter introduces concepts, notations, and algorithms relevant for this thesis. The survey starts with a brief summary of basic notations and concepts in probability theory in Section 3.1. In Section 3.2, a compact introduction of probabilistic graphical models is given. For a broader introduction with lots of explaining examples and exercises, the reader is kindly referred to the book of Koller and Friedman [61]. Relevant discrete inference methods are introduced in Section 3.3. The last two parts of this chapter are dedicated to an introduction to stochastic methods for inference with continuous variables in Section 3.4 and its application on probabilistic graphical models with a widely used approach named particle belief propagation in Section 3.5.

3.1 Notations

Let capital letters (for instance X) denote *random variables* over *probability distributions* $P(X)$. A random variable associates each outcome of a random process to a value (or *configuration*), denoted with small letters (e.g. x), from a (measurable) *state space* $x \in \mathcal{X}$. Let $p(x)$ denote either a *probability mass function* when X is discrete, or a *probability density function* when X is continuous, respectively.

Let $X = (X_1, ...X_N)$ be a *multivariate random variable*, where X_n are (scalar) random variables. Then, $P(X_1, ...X_N)$ is called the *joint probability distribution.* The distribution $P(X_n)$ over a single element (or a subset) of X is called a *marginal distribution over* X_n and is defined as

$$P(X_n) = P(\Omega_1, \ldots, \Omega_{n-1}, X_n, \Omega_{n+1}, \ldots, \Omega_N), \tag{3.1}$$

where Ω_n is the certain event (i.e. $P(\Omega_n) = 1$) for the n-th random variable. The corresponding probability mass function for a discrete random variable is

$$p(x_n) = \sum_{m \neq n, x_m \in \mathcal{X}_m} p(x_1, ..., x_N). \tag{3.2}$$

For continuous random variables, the sum is replaced with an integral. To highlight the use of a marginal distribution (with respect to a joint distribution), the symbol ν is used instead of p for marginal functions.

Another important type of distributions are conditional probability distributions. The *conditional probability* of X given Y, denoted by $p(x \mid y)$ is the probability of X when the value of Y is known. Conditional and joint probabilities are related to each other via the *chain rule*:

$$p(x, y) = p(x \mid y)p(y). \tag{3.3}$$

The final and most important concept in context of the following chapters is *conditional independence* of random variables. This is the main attribute of joint

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

probabilities which will be exploited to build a powerful modeling language and to construct efficient inference and learning algorithms.

A random variable X is *independent* of another random variable Y, if the conditional probability $p(x \mid y)$ does not depend on Y, that is $p(x \mid y) = p(x)$ (assuming that y *can* occur, i.e. $p(y) > 0$; otherwise, independence is trivially fulfilled). So, independently of which y is chosen, the probability $p(x \mid y)$ does not change. Equivalent to that is the following definition. Two (sets of) random variables X and Y are said to be independent, if and only if

$$p(x,y) = p(x)p(y) \qquad \forall x \in \mathcal{X}, y \in \mathcal{Y}. \tag{3.4}$$

We write $X \perp\!\!\!\perp Y$ as shorthand for X is independent of Y. This definition can be further extended through conditioning on a third (set of) random variable(s). Two (sets of) random variables X and Y are said to be *conditional independent* of Z, if and only if

$$p(x,y \mid z) = p(x \mid z)p(y \mid z) \qquad \forall x \in \mathcal{X}, y \in \mathcal{Y}, z \in \mathcal{Z}. \tag{3.5}$$

We write $X \perp\!\!\!\perp Y \mid Z$ as shorthand for X is conditionally independent of Y given Z.

3.2 Probabilistic Graphical Models

First, an intuitive introduction to probabilistic graphical models (PGMs) is given and advantages in using PGMs are illustrated. This is followed by a formal definition of different PGM representations with its corresponding advantages and weaknesses.

Intuitively said, a PGM is the marriage between probability theory and graph theory. The goal is to visualize certain properties of a probability distribution over a large set of random variables with entities known from graph theory such as nodes and edges. This approach can be used to *analyze* a given family of probability distributions simply by taking a look on its corresponding graph representation. On the other hand, PGMs provide a very powerful *modeling* framework to define certain assumptions on the underlying probability distribution. Furthermore, PGMs provide a conceptional framework to define appropriate data structures for constructing efficient inference and learning algorithms. A basic tool to put structure on probability distributions are (conditonal) independence assumptions, as introduced in Sect. 3.1.

In the literature, there exist a variety of PGM representations for characterization and visualization of probability distributions with an underlying structure. Each representation has its strengths and weaknesses in visualizing various properties of the distributions. The most common used representations for PGMs include Bayesian networks, Markov networks (or Markov random fields) and factor graphs. Bayes networks and Markov networks represent different families of independence assumptions. Bayes networks are the tool of choice for modeling natural directionality (e.g.

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

derived from causality considerations). Symmetrical interactions cannot be adequately modeled in Bayes networks. This is the domain of Markov networks. Factor graphs provide a unified and finer-grained view on the factorization of probability distributions than Markov networks and Bayes networks. They are preferably used in construction of algorithms for inference and parameter learning.

For a detailed discussion on various modeling aspects and theoretical properties of various PGMs and its relations to each other, the interested reader is kindly referred to [61].

3.2.1 Bayes Network

A Bayes network is a probabilistic graphical model where the underlying structure is a directed acyclic graph (DAG) $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with nodes (or *vertices*) $\mathcal{V}$ and directed edges $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$. A directed edge from node s to node t is denoted as $(s, t) \in \mathcal{E}$ or equivalently $(s \to t) \in \mathcal{E}$. A DAG must not contain cycles. This property suggests the use of parent-child relationships. That is, the set of parents $\mathrm{Pa}(s)$ of a vertex s is defined as $\mathrm{Pa}(s) = \{t \in \mathcal{V} \mid (t \to s) \in \mathcal{E}\}$.

A Bayes network is defined as the pair $(\mathcal{G}, P)$, where $\mathcal{G}$ is a DAG and P a joint probability distribution over $|\mathcal{V}|$ random variables. Each vertex $s \in \mathcal{V}$ is associated to a random variable X_s and P factorizes into

$$P(X_1, ..., X_{|\mathcal{V}|}) = \prod_{s \in \mathcal{V}} P(X_s \mid X_{\mathrm{Pa}(s)}). \tag{3.6}$$

3.2.2 Markov Random Field

A Markov network, or Markov random field (MRF), is defined over a graph $\mathcal{G} = (\mathcal{V},\mathcal{E})$ with a set of vertices $\mathcal{V}$ and a set of undirected edges $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$ associated with a (strictly positive) probability distribution P. The vertex $s \in \mathcal{V}$ is connected by an edge with vertex $t \in \mathcal{V}$ if, and only if $(s,t) \in \mathcal{E}$. Edges are undirected, such that $(s,t) \in \mathcal{E} \Leftrightarrow (t,s) \in \mathcal{E}$. The definition of a *neighborhood system* $\mathcal{N}$ on graph $\mathcal{G}$ is useful, where $\mathcal{N}_s = \{t \in \mathcal{V} \mid (s,t) \in \mathcal{E}\}$ is the set of neighbors to node s and $\mathcal{N} = \{\mathcal{N}_s\}_{s \in \mathcal{V}}$. A fully connected (unordered) subset $c \subset \mathcal{V}$ is called a *clique*. The set $\mathcal{C}$ is the set of all cliques in graph $\mathcal{G}$. Figure 3.1a shows an MRF over five vertices. Vertex 1 is connected with 2 and 5 and thus $\mathcal{N}_1 = \{2, 5\}$. Figure 3.1c shows some cliques of the MRF in Figure 3.1a.

Each vertex $s \in \mathcal{V}$ is associated with a random variable X_s. The realization of a random variable X_s is denoted with x_s and $\mathcal{X}_s$ is its state space such that $x_s \in \mathcal{X}_s$. The joint realization, or configuration, of $\mathcal{G}$ is $\mathbf{x} = (x_1, ..., x_{|\mathcal{V}|})^\mathsf{T}$, and the joint state space is the cartesian product $\mathcal{X} = \prod_{s \in \mathcal{V}} \mathcal{X}_s$.

The distribution P is called a Markov random field with respect to graph $\mathcal{G}$ if for

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

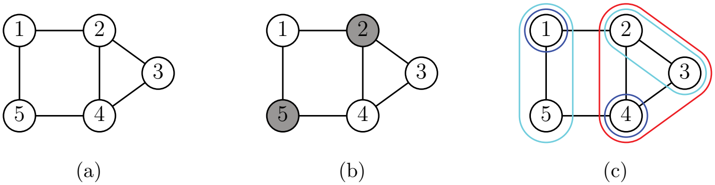


Figure 3.1: (a) An example Markov network with vertices $\mathcal{V} = \{1,2,3,4,5\}$ and edges $\mathcal{E} = \{(1,2),(1,5),(2,3),(2,4),(3,4),(4,5)\}$. (b) Illustration of Markov property and conditional independence in Eq. (3.8) with $s = 1$, $\mathcal{N}_s = \{2,5\}$, and $\mathcal{V} \setminus \mathcal{N}_s \cup \{s\} = \{3,4\}$. (c) Some (not exhaustive) 1-cliques (blue), 2-cliques (cyan), and 3-cliques (red) of the Markov network from (a). The set $\{1,2,4,5\}$ does not form a 4-clique, since $(1,4) \notin \mathcal{E}$ and $(2,5) \notin \mathcal{E}$.

all $s \in \mathcal{V}$ and for all $\mathbf{x} \in \mathcal{X}$:

$$p(x_s \mid \mathbf{x}_{\mathcal{V}\setminus\{s\}}) = p(x_s \mid \mathbf{x}_{\mathcal{N}_s}). \tag{3.7}$$

The conditional probability of any random variable given all other random variables is the same as conditioning that random variable only on its direct neighbors. That is, conditional probabilities can be fully specified by only considering directly neighbored variables. All other variables do not have an influence on the conditional probability. This *locality* principle or *Markov property* is what makes MRFs a very powerful tool. Figure 3.1b illustrates this on the example network of Figure 3.1a. Here, the conditional probability of x_1 given the rest is considered. Variables x_3 and x_4 have no influence and can thus take any value since they are "blocked" from x_1 through x_2 and x_5 (dark-gray vertices).

If $p(\mathbf{x}) > 0 \; \forall \mathbf{x} \in \mathcal{X}$, then the following equivalence holds:

$$p(x_s \mid \mathbf{x}_{\mathcal{V}\setminus\{s\}}) = p(x_s \mid \mathbf{x}_{\mathcal{N}_s}) \quad \Leftrightarrow \quad X_s \perp\!\!\!\perp X_{\mathcal{V}\setminus\mathcal{N}_s\cup\{s\}} \mid X_{\mathcal{N}_s}. \tag{3.8}$$

The property $X_s \perp\!\!\!\perp X_{\mathcal{V}\setminus\mathcal{N}_s\cup\{s\}} \mid X_{\mathcal{N}_s}$ is called the *local Markov property*. That is, a MRF can be specified via conditional independence assumptions (cf. Sect. 3.1). In the example of Fig. 3.1b it holds $X_1 \perp\!\!\!\perp X_3,X_4 \mid X_2,X_5$. The set $X_{\mathcal{N}_s}$ is also known as the *Markov blanket*.

Intuitively said, a direct dependence of random variables is visualized by connecting them via edges. The *absence* of an edge hence indicates conditional independence as already illustrated in Figure 3.1b. Note that this does not exclude long-range dependencies. For example, in Figure 3.1a variable X_1 can influence variable X_3 although they are not connected through a direct edge but through at least one *path* (for example the path 1–2–3). On the other hand, when the variables X_2 and

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

X_5 are observed, the dependence disappears since there is no path which does not contain the observed variables.

The definition in Eq. 3.7 already suggests that the joint probability function of an MRF can be decomposed into a set of local functions. Now the question arises how this decomposition looks like? The answer is given by taking a look on *Gibbs distributions*.

Gibbs Distribution A probability distribution is called a *Gibbs distribution* over a graph $\mathcal{G}$ if it is a normalized factorization over the cliques of $\mathcal{G}$:

$$p(\mathbf{x}) = \frac{1}{Z} \prod_{c \in \mathcal{C}} \phi_c(x_c) \tag{3.9}$$

with the normalizer or *partition function* $Z = \sum_{\mathbf{x}} \prod_{c \in \mathcal{C}} \phi_c(x_c)$. The functions $\phi_c(x_c)$ must be positive and are called *clique potential functions* (or short *clique potentials*). Note that the ordering of vertices within a clique c is irrelevant, such that $\phi_{c1,c_2,\dots,c_{|c|}}(x_{c_1},x_{c_2},\dots,x_{c_{|c|}}) = \phi_{c2,c_1,\dots,c_{|c|}}(x_{c_2},x_{c_1},\dots,x_{c_{|c|}}) = \dots$ is required. Furthermore, the potential functions are not required to be valid probabilities (i.e. they are not normalized). It is convenient to rewrite a Gibbs distribution in log-linear form

$$p(\mathbf{x}) \propto \exp[-E(\mathbf{x})] \qquad\qquad E(\mathbf{x}) = \sum_{c \in \mathcal{C}} \psi_c(x_c) \tag{3.10}$$

where $E(\mathbf{x})$ is called an *energy function* and the *clique energies* are defined through the clique potentials via $\psi_c(x_c) = -\log \phi_c(x_c)$. This representation is convenient as the positivity constraint $\phi_c(x_c) > 0$ is fulfilled for all real valued (and finite) $\psi_c(x_c) \in (-\infty, \infty)$.

MRFs and Gibbs distributions are related to each other by the Hammersley-Clifford theorem which states that each MRF with strictly positive P can be expressed by a Gibbs distribution and vice versa. Recall that MRFs are characterized by conditional independence (cf. Eq. (3.7)) and a Gibbs distribution is characterized by its function factorization (cf. Eq. (3.9)). Hence, the Hammersley-Clifford theorem guarantees (for strictly positive P) a factorization of an MRF into local factors and each factorization of a probability distribution can be expressed as an MRF.

Pairwise Markov Random Field An important special case of MRFs are *pairwise MRFs*. If the graph $\mathcal{G}$ only consists of *singleton* and *pairwise* cliques, i.e. $|c| \leq 2$, then the factorization can be rearranged to the following special form

$$p(\mathbf{x}) \propto \prod_{s \in \mathcal{V}} \phi_s(x_s) \prod_{(s,t) \in \mathcal{E}} \phi_{s,t}(x_s, x_t) \tag{3.11}$$

with singleton (or unary) potentials $\phi_s(x_s)$ and pairwise (or binary) potentials $\phi_{s,t}(x_s, x_t)$.

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

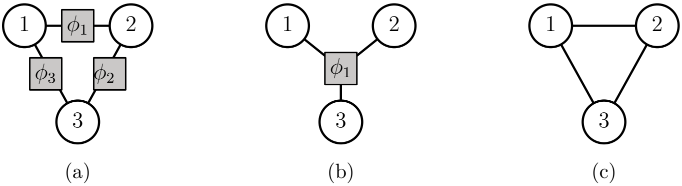


Figure 3.2: (a) Factor graph of distribution $p(x_1, x_2, x_3) \propto \phi_1(x_1, x_2)\phi_2(x_2,x_3)\phi_1(x_3, x_1)$. (b) Factor graph of distribution $p(x_1, x_2, x_3) \propto \phi_1(x_1, x_2, x_3)$. (c) Same Markov network representation for both factor graphs in (a) and (b).

Consider again the example Markov network in Figure 3.1a. Assume that this network is a pairwise MRF. Then the vertices of the graph correspond exactly to the unary potentials and the edges of the graph correspond exactly to the pairwise potentials. This one-to-one correspondence of the graph entities to the MRF factorization makes the Markov network notation a very attractive tool for visualizing pairwise MRFs.

Note that when dealing with higher-order graphical models, the Markov network notation is no longer expressive enough to conclude about the (intended) factorization of the underlying probability distribution. The factor graph notation introduced in the following Section provides a finer-grained representation.

3.2.3 Factor Graph

Factor graphs are a convenient way for representing the factorization of functions such as Gibbs distributions (cf. Eq. (3.9)). Furthermore, factor graphs allow a unified encoding of the diverse PGM representations such as Markov networks or Bayes networks. This allows, e.g., for a very generic implementation of diverse inference and learning algorithms [57].

A factor graph (cf. [72]) is a graph $\mathcal{G} = (\mathcal{V} \cup \mathcal{F}, \mathcal{E})$ with two disjoint sets of vertices – random variable vertices $\mathcal{V}$ and factor vertices $\mathcal{F}$ with $\mathcal{V} \cap \mathcal{F} = \emptyset$ – and a set of edges $\mathcal{E} \subset \mathcal{V} \times \mathcal{F}$ connecting random variable vertices with factor vertices. Random variable vertices are analogously to Markov networks drawn as circles. Factor vertices are drawn as dark-gray filled rectangles. Note that edges connecting random variables with random variables or factors with factors are not allowed and thus a factor graph is always a bipartite graph.

As already mentioned before, Markov networks have some inherent ambiguity in defining the clique factorization. Factor graphs are finer-grained. Figure 3.2 illustrates this (dis-)ambiguity on a minimal example.

For efficient inference in probabilistic graphical models as will be discussed in

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

the following sections of this chapter, an appropriate representation of factors is of fundamental importance. Therefore, before introducing the inference methods, some common representations for factors over discrete random variables and over continuous random variables will be discussed.

Discrete Factor Representation If the state space is finite and discrete, a typical representation for factors is a *lookup table*. Let the state space be $\mathcal{X}_s = \{{x_s}^1, ..., {x_s}^{L_s}\}$, where $L_s \leq L$ is the number of states for vertex s (with at most L states per vertex). Then a factor $\phi_c(\mathbf{x}_c)$, with associated variables $s \in c = \{c_1, ..., c_{|c|}\}$ can be implemented as a multi-dimensional array

$$\phi_c(\mathbf{x}_c) = \phi_c[x_{c_1}, ..., x_{c_K}] \tag{3.12}$$

with $K = |c|$. It is clear that the memory consumption per factor is $\mathcal{O}(L^K)$, i.e. exponential in the clique size. For pairwise MRFs, the memory consumption for the complete graph then is $\mathcal{O}(L|\mathcal{V}| + L^2|\mathcal{E}|)$.

Often, more memory (and thus computationally) efficient parametrizations can be found by exploiting some knowledge about the factor potentials. One prominent example is the Potts model:

$$\phi_{s,t}(x_s, x_t) = \begin{cases} \theta_{\text{eq}} & \text{if } x_s = x_t \\ \theta_{\text{ineq}} & \text{if } x_s \neq x_t. \end{cases} \tag{3.13}$$

Here, the parameter set contains only two elements $\theta = (\theta_{\text{eq}}, \theta_{\text{ineq}})$, regardless of the number of labels.

An important class, especially for parameter learning, are log-linear models:

$$\psi_c(x_c) = \sum_{i=1}^{k} w_i \cdot f_{c,i}(x_c) = \theta_c^\mathsf{T} f_c(x_c), \tag{3.14}$$

where $\theta_c = (w_1, ..., w_k)$ and f_c is a feature function. It is important to note that the linearity is with respect to the parameters and not to the random variables, as the feature function f_c can introduce arbitrary non-linearity. The linearity with respect to θ is very advantageous when it comes to parameter learning, since the (features of the) training samples can be summarized into *sufficient statistics*, which are linearly weighted by the parameters.

Continuous Factor Representation One can divide the factor representations for continuous state spaces in two groups. The parametrized and the non-parametrized representations. While there exists a great variety of parametrized function families, only a very limited subset is suitable for enabling efficient inference in factor graphs. When performing inference, it is often necessary to *summarize* the influence of a whole group of factors into a single (virtual) factor. Therefore, the

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

factor parametrization should be invariant with respect to such operations. One of such very rare families is the Gaussian distribution:

$$\phi_c(x_c) \propto \exp[-0.5(\mathbf{x}_c - \boldsymbol{\mu}_c)^\mathsf{T}\Sigma_c^{-1}(\mathbf{x}_c - \boldsymbol{\mu}_c)] \tag{3.15}$$

with parameters $\theta_c = (\boldsymbol{\mu}_c, \Sigma_c)$, where $\boldsymbol{\mu}_c$ is the mean vector and Σ_c is the (symmetric positive definite) covariance matrix. If all factors are Gaussian distributed, then the joint distribution is also Gaussian. Consider, for example, the computation of the marginal distribution $P(X_n)$ in Eq. 3.1 from the joint distribution $P(X_1, \ldots, X_N)$. One can show that the marginal distribution of a Gaussian joint distribution is again a Gaussian distribution [15].

In general, complex or multi-modal distributions cannot be approximated well by (uni-modal) Gaussian distributions. One way to increase the expressiveness is to resort to *hybrid models*. Here, the joint distribution uses both continuous and discrete random variables. The factor potentials can then be modeled as mixtures of Gaussians or logistic functions [61]. In Gaussian mixture models, the discrete variables can be interpreted as acting like *switches*; that is, there is an own parameter set for each discrete state:

$$\phi_{c,t}(\mathbf{x}_c, y_t) \propto \exp[-0.5(\mathbf{x}_c - \boldsymbol{\mu}_c^{y_t})^\mathsf{T}(\Sigma_c^{y_t})^{-1}(\mathbf{x}_c - \boldsymbol{\mu}_c^{y_t})]. \tag{3.16}$$

Non-parametrized representations can be applied to a much broader set of distributions. Here, the factor potentials can take any form. The summary operations are approximated using non-parametric representations such as Gaussian mixture models [107, 106] or Monte Carlo simulation [50, 67].

In this thesis, the latter approach is followed. Here, the continuous state space is represented using a finite set of particles $\{x_s^{(i)}\}_{i=1}^{p} = \mathcal{P} \subset \mathcal{X}$. The factors are only evaluated at the position of the particles. Thus, only the values at these positions have to be stored. This can be handled analogously to the discrete case using lookup tables (cf. Eq. (3.12)).

3.3 Inference in Probabilistic Graphical Models

The goal of this thesis is to apply graphical models in various computer vision applications. Inference or reasoning in probabilistic graphical models is the process of answering queries of interest using joint probability distributions [61]. One of the most important inference tasks to the computer vision and machine learning community is maximum a posteriori (MAP) inference. MAP inference is the task of inferring the most probable joint configuration $\mathbf{x}^*$ from the joint posterior distribution $p(\mathbf{x} \mid \mathbf{d})$ conditioned on observed data $\mathbf{d}$ as

$$\mathbf{x}^* = \arg\max_{\mathbf{x}\in\mathcal{X}} p(\mathbf{x} \mid \mathbf{d}). \tag{3.17}$$

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

According to the chain rule, the posterior can be rewritten as $p(\mathbf{x} \mid \mathbf{d}) = p(\mathbf{x}, \mathbf{d})/p(\mathbf{d})$. Note that the evidence $p(\mathbf{d})$ is a constant and thus can be omitted:

$$\mathbf{x}^* = \arg\max_{\mathbf{x}\in\mathcal{X}} \frac{p(\mathbf{x}, \mathbf{d})}{p(\mathbf{d})} = \arg\max_{\mathbf{x}\in\mathcal{X}} p(\mathbf{x}, \mathbf{d}). \tag{3.18}$$

Hence, the posterior needs only be available up to a scaling factor. This considerably reduces computational complexity since the normalization factor $p(\mathbf{d})$ is in general very hard to compute.

In computer vision, many tasks can be reformulated as MAP inference problems, since one often seeks only for the most probable solution.

A large class of MAP inference algorithms based on the so called *message passing* principle works in two steps: First, so-called *max-marginals* are computed:

$$\mu_s(x_s) = \max_{\mathbf{y}\in\mathcal{X}|y_s=x_s} p(\mathbf{y}). \tag{3.19}$$

It can be interpreted as the unnormalized probability of the most likely joint assignment consistent with x_s [61]. It has a close relationship to the marginal distribution as introduced in Section 3.1, Eq. (3.1). Its formal connection to the marginal function is obvious: the summation operator is simply replaced by a max-operator. In a second step, the max-marginals are used to construct the MAP configuration. This step is known as *decoding*.

The detour over max-marginals has some advantages compared to direct MAP inference approaches. In case of tree structured graphical models, the max-marginals can be efficiently calculated with methods based on dynamic programming known as *max-product belief propagation* which is introduced in the following Section. For non-tree structured graphs (loopy graphs), the same dynamic programming approach leads to a method called *max-product loopy belief propagation*. Although this method is not exact for loopy graphs, it often yields reasonably well approximates but with very weak convergence guarantees. Methods based on dual formulations have better convergence guarantees and additionally provide upper bounds. These are introduced in Section 3.3.3.

A second advantage is that max-marginals are itself (non-normalized) distributions and thus provide richer information about the target distribution than the “single shot” MAP estimate. This can be exploited to enable efficient inference over continuous random variables using stochastic search methods. Section 3.4 gives an overview of stochastic search methods and in Section 3.5 and Chapter 4 these methods are applied on the task of MAP inference.

In the following section 3.3.1, a very basic method for calculating the (max-) marginals and MAP in tree structured graphs is introduced. Afterwards in Secs. 3.3.2–3.3.3, (approximate) methods are introduced which also work in loopy graphical models.

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

3.3.1 Max-Product Belief Propagation

The max-product belief propagation (MP-BP) algorithm is a method to compute the max-marginals $\mu_s(x_s)$ in tree structured graphs.

Let us start with a simple example. Consider the joint density function $p(a, b, c, d) = \phi_a(a)\phi_{a,b}(a, b)\phi_{b,c}(b, c)\phi_{b,d}(b, d)$. The goal is to compute the max-marginals $\mu_a(a)$, $\mu_b(b)$, $\mu_c(c)$, and $\mu_d(d)$. A straight forward approach to compute, e.g., $\mu_a(a)$ is to directly apply its definition (cf. Eq. (3.19))

$$\mu_a(a) = \max_{b,c,d} p(a, b, c, d) = \max_{b,c,d} \phi_a(a)\phi_{a,b}(a, b)\phi_{b,c}(b, c)\phi_{b,d}(b, d).$$

One can observe that maximization over, say, variable d does only involve the factor $\phi_{b,d}(b, d)$. The rest is constant wrt. d. Thus, the max-operation over d can be "pulled out":

$$\mu_a(a) = \max_{b,c} \phi_a(a)\phi_{a,b}(a, b)\phi_{b,c}(b, c) \underbrace{\max_{d} \phi_{b,d}(b, d)}_{m_{d\to b}(b)}.$$

The new factor $m_{d\to b}(b) = \max_d \phi_{b,d}(b, d)$ is not dependent on d anymore. The variable d is *eliminated*, hence the algorithm name *variable elimination.* Further elimination of c and b leads to

$$\mu_a(a) = \phi_a(a) \underbrace{\max_{b}[\phi_{a,b}(a, b) \underbrace{\max_{c} \phi_{b,c}(b, c)}_{\text{1a: } m_{c\to b}(b)} \underbrace{\max_{d} \phi_{b,d}(b, d)}_{\text{1b: } m_{d\to b}(b)}]}_{\text{2: } m_{b\to a}(a)}.$$

Note the elimination order 1a, 1b, 2. Steps 1a and 1b are done independently of each other and could be switched in order as well. Thus, the visiting schedule is a partial ordering. The same procedure can be repeated for marginalizing over, e.g., variable d:

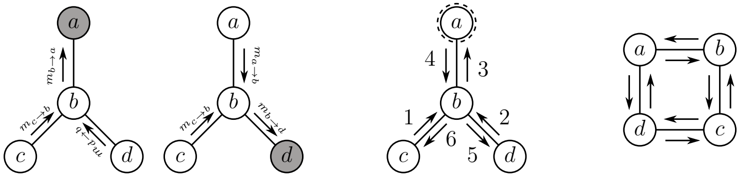
$$\mu_d(d) = \underbrace{\max_{b}[\underbrace{\max_{a} \phi_a(a)\phi_{a,b}(a, b)}_{\text{1a: } m_{a\to b}(b)} \underbrace{\max_{c} \phi_{b,c}(b, c)}_{\text{1b: } m_{c\to b}(b)} \phi_{b,d}(b, d)]}_{\text{2: } m_{b\to d}(d)}.$$

Figure 3.3a summarizes the variable elimination steps for $\mu_a(a)$ and $\mu_d(d)$.

An important observation is that the calculation of $m_{d\to b}(b)$ is actually done twice (one time for calculating $\mu_a(a)$ and a second time for calculating $\mu_d(d)$). Another observation is that the *elimination order* can drastically influence the computational effort. If variable b is eliminated first, then the maximization must be performed over the remaining three variables (b,c,d in case of $\mu_a(a)$, and a,b,c in case of $\mu_d(d)$).

The MP-BP algorithm can be interpreted as a parallelization of the variable elimination algorithm where intermediate computation steps are *cached.* Caching of

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.



(a) (b) (c)

Figure 3.3: (a) Two example instances of the variable elimination algorithm for calculating marginals of vertices a and d, respectively, and its corresponding message evaluation orders. (b) Belief propagation message passing scheme on a tree. The dashed circle indicates the root node. The numbers indicate the visiting order for message computation (cf. Eq. (3.20) and Alg. 1). (c) Loopy Belief propagation message passing scheme.

these intermediate results leads to significant computational savings since these values would be otherwise computed multiple times. The intermediate results $m_{s\to t}$ are called *messages* (from node s to node t). The computational effort for calculating the max-marginals with MP-BP is $\mathcal{O}(2|\mathcal{E}|)$, whereas with the naive variable elimination approach (without caching) it would be $\mathcal{O}(|\mathcal{V}|\cdot|\mathcal{E}|)$. Figure 3.3b depicts an instance of MP-BP for the above toy example.

In the following, the exemplary procedure above is formulated for general tree structured graphical models.

Max-Marginals: Max-Product Message Passing

Recall, that graph $\mathcal{G}$ must form a tree (i.e. $\mathcal{G}$ does not contain loops). Then, there exists a partial ordering $(r \to t) \prec (t \to s)$ (a *visiting schedule*) which defines a valid order in which to compute the messages as follows:

$$m_{t\to s}(x_s) \propto \max_{x_t\in\mathcal{X}_t}[\phi_t(x_t)\phi_{s,t}(x_s,x_t)\prod_{r\in\mathcal{N}_t\setminus\{s\}} m_{r\to t}(x_t)]. \tag{3.20}$$

A partial ordering is valid, if all *ingoing* messages wrt. some vertex t (i.e. all messages $\{m_{s\to t}(x_t)\}_s$) are calculated prior to all *outgoing* messages wrt. t (messages $\{m_{t\to s}(x_s)\}_s$). A valid visiting schedule can be constructed by using the following scheme: First, choose an arbitrary node as the tree's *root*. Then, compute all messages pointing towards the root node, beginning at the leafs and ending at the root. Third, compute all messages pointing away from the root node, beginning at the root and ending at the leafs. This schedule is depicted in Fig. 3.3b.

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

Algorithm 1 Max-Product Belief Propagation

```
**Input:** Gibbs factorization $\{\phi_s\}_{s\in\mathcal{V}}$, $\{\phi_{st}\}_{(s,t)\in\mathcal{E}}$, and visiting schedule $\mathcal{S}$.
**Ensure:** Max-Marginals $\{\mu_s\}_{s\in\mathcal{V}}$, MAP configuration $\mathbf{x}^* \in \arg\max_{\mathbf{x}\in\mathcal{X}} p(\mathbf{x})$
for $(t \to s) \in \mathcal{S}$ in visiting order do
    Compute message $m_{t\to s}(x_s) \propto \max_{x_t\in\mathcal{X}_t} [\phi_t(x_t)\phi_{s,t}(x_s,x_t) \prod_{r\in\mathcal{N}_t\setminus\{s\}} m_{r\to t}(x_t)]$
end for
for each node $s$ do
    Compute belief $\mu_s(x_s) \propto \phi_s(x_s) \prod_{t\in\mathcal{N}_s} m_{t\to s}(x_s)$
end for
Decode $x_s^* \in \arg\max_{x_s\in\mathcal{X}_s} \mu_s(x_s)$
```

Finally, the max-marginals are calculated as the product of the unary potential and all incoming messages:

$$\mu_s(x_s) \propto \phi_s(x_s) \prod_{t\in\mathcal{N}_s} m_{t\to s}(x_s). \tag{3.21}$$

MAP Decoding from Max-Marginals

Decoding is the task of inferring a MAP configuration $\mathbf{x}^* \in \arg\max_{\mathbf{x}\in\mathcal{X}} p(\mathbf{x})$ from the max-marginals. One can show that if and only if the max-marginals have unique maximizers then a unique MAP configuration exists and the decoding process reduces to finding local maximizers of the max-marginals [61]:

$$x_s^* = \arg\max_{x_s\in\mathcal{X}_s} \mu_s(x_s). \tag{3.22}$$

Note, that this is not necessarily the case. For instance, consider a discrete distribution with two random variables with state space $\mathcal{X} = \{0,1\}$. Assume that the unary potentials are non-informative with $\phi_s(0) = \phi_s(1) = 0.5$ and the pairwise potential connecting both variables prefers equal states, but is otherwise completely symmetric: $\phi_{s,t}(0,0) = \phi_{s,t}(1,1) = 0.9$, $\phi_{s,t}(0,1) = \phi_{s,t}(1,0) = 0.1$. Then, the max-marginals are $\mu_s(0) = \mu_s(1) = 0.5$. This suggest that any joint configuration $(x_1,x_2) \in \{(0,0),(0,1),(1,0),(1,1)\}$ would maximize the joint probability. This is obviously not the case as only the two joint configurations $\{(0,0),(1,1)\}$ are solutions of the original MAP problem. To resolve such cases correctly, a backtracking approach using, e.g., dynamic programming must be applied [116]. Algorithm 1 summarizes the MP-BP method.

3.3.2 Max-Product Loopy Belief Propagation

If the graph $\mathcal{G}$ is not a tree then the recursive message update in Eq. (3.20) cannot be applied directly. That is because this recursion contains *circular* dependencies.

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

E.g. consider the loopy graph in Fig. 3.3c. In order to compute $m_{a\to b}$, the message from node a to node b, one has to compute $m_{d\to a}$ first. For computing this message in turn, the message $m_{c\to d}$ has to be available and prior to this, message $m_{c\to b}$ has to be known. But message $m_{c\to b}$ again depends on message $m_{a\to b}$, which is our desired unknown.

Approaches to workaround this dilemma resort to *iterative* computation of the messages. In the initialization step, all messages are typically set to the non-informative, uniform distribution $m_{t\to s}(x_s) \equiv 1$. In subsequent iteration steps, the messages are updated similar to Eq. (3.20) using a predefined message passing schedule. But instead of using the true (possibly unknown) incoming messages $m_{r\to t}(x_t)$ for computing $m_{t\to s}(x_s)$, the (approximate) messages of the previous iteration step are used instead.

Now, the hope is that after sufficient iterations an equilibrium is reached, also called a fixpoint. In general, this heuristic scheme is not guaranteed to converge. And even if it converges to a fixpoint, the computed quantities in Eq.(3.21) are no longer guaranteed to match the true max-marginals. Furthermore, convergence and speed of convergence highly depend on the chosen message passing schedule. Selecting optimal message passing schedules is an open issue.

The (max-)marginals inferred using (max-product) loopy belief propagation (BP) or other approximate (max-product) message passing methods are usually referred to as *pseudo-(max-)marginals*.

Practical Considerations Implementation of the max-product (loopy) belief propagation algorithms is fairly easy and straight forward. However, some practical issues have to be considered in order to ensure correct behavior [89].

It is advantageous to implement the inference algorithms in negative log-space in order to increase numerical stability. That is, $M_{t\to s}(x_s) = -\log[m_{t\to s}(x_s)]$, $B_s(x_s) = -\log[\mu_s(x_s)]$, and $N_s(x_s) = -\log[\nu_s(x_s)]$. Applying the negative logarithm transforms products to sums and max-operators to min-operators. After some rearrangement, the max-product message-passing formulas transform to *min-sum message-passing* as follows:

$$M_{t\to s}(x_s) = \min_{x_t \in \mathcal{X}_t} [\psi_{s,t}(x_s,x_t) + B_t(x_t) - M_{s\to t}(x_t)] \tag{3.23}$$

$$B_s(x_s) = \psi_s(x_s) + \sum_{t \in \mathcal{N}_s} M_{t\to s}(x_s). \tag{3.24}$$

A further issue is the unbounded growth of the message values after each message passing operation (cf. Eq. (3.20)). The messages are normalized as follows $M_{t\to s}(x_s) \leftarrow M_{t\to s}(x_s) - \min_{x'_s} M_{t\to s}(x'_s)$. Algorithm 2 summarizes the loopy min-sum message passing implementation.

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

Algorithm 2 Loopy Min-Sum Algorithm

```
Input: Factor potentials ${\psi_s}_{s \in \mathcal{V}}$, ${\psi_{st}}_{(s,t) \in \mathcal{E}}$
Ensure: Negative log pseudo max-marginals ${B_s}_{s \in \mathcal{V}}$, MAP estimate $\mathbf{x}^*$
Initialize the messages $M^0_{t \to s}(x_s) = 0$ and log disbelief $B^0_s(x_s) = 0 \; \forall s,t$
for BP iteration $n = 1$ to $N$ do
    for each node $s$ do
        for each $t \in \mathcal{N}_s$ do
            $M^n_{t \to s}(x_s) = \min_{x_t \in \mathcal{X}_t} [\psi_{s,t}(x_s,x_t) + B^{n-1}_t(x_t) - M^{n-1}_{s \to t}(x_t)]$
        end for
        $B^n_s(x_s) = \psi_s(x_s) + \sum_{t \in \mathcal{N}_s} M^n_{t \to s}(x_s)$
    end for
    Normalize messages: $M^n_{t \to s}(x_s) := M^n_{t \to s}(x_s) - \min_{x'_s} M^n_{t \to s}(x'_s) \; \forall s,t$
    Normalize beliefs: $B^n_s(x_s) := B^n_s(x_s) - \min_{x'_s} B^n_s(x'_s) \; \forall s$
end for
Decode $x^*_s \in \arg\min_{x_s \in \mathcal{X}_s} B^N_s(x_s)$
```

3.3.3 Dual Methods

The previously introduced loopy belief propagation algorithm can be seen as a heuristic method in case of loopy graphs. It does not have a convergence guarantee and even if a fix-point is reached, optimality is not guaranteed. The methods which are presented in the following Sections are based on Lagrangian dual formulations of Eqs. (3.19) and (3.17), respectively. Due to its superior properties both in theory and practical application, these methods will be used in the consecutive chapters.

Tree-Reweighted Max-Product Belief Propagation

The tree-reweighted belief propagation (TRBP) algorithm of [116] is a modification of the MP-BP method. It originates from the idea of decomposing the original graphical model into arbitrary sets of spanning trees $\mathcal{T}$. The MAP problem is then (approximately) solved by using convex combinations of these spanning trees. The resulting iterative optimization method turns out to have surprising similarity to the ordinary BP method in section 3.3.1:

$$m_{t \to s}(x_s) \propto \max_{x_t \in \mathcal{X}_t} \phi_t(x_t) \phi_{s,t}(x_s, x_t)^{\frac{1}{\rho_{st}}} \frac{\prod_{r \in \mathcal{N}_t \setminus \{s\}} m_{r \to t}(x_t)^{\rho_{st}}}{m_{s \to t}(x_t)^{1-\rho_{st}}} \tag{3.25}$$

$$\mu_s(x_s) \propto \phi_s(x_s) \prod_{t \in \mathcal{N}_s} m_{t \to s}(x_s)^{\rho_{ts}}. \tag{3.26}$$

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

The basic message passing mechanism is preserved. The TRBP message propagation and pseudo-max-marginal formulas include weighting factors ρ_{st} for each edge $(s,t) \in \mathcal{E}$ (note, again, the symmetry $\rho_{st} = \rho_{ts}$). Additionally, unlike to ordinary BP, the TRBP message propagation from node s to node t (denoted as $s \rightarrow t$) also depends on the reverse message $t \rightarrow s$. The weights ρ_{st} are the probabilities of the edges (s,t) being part of a tree τ in the set of all spanning trees $\mathcal{T}$. Their values can be computed by first sampling a random set of spanning trees $\mathcal{T}$ from graph $\mathcal{G}$ such that each edge is covered at least once. The weight ρ_{st} is then the number of spanning trees covering the corresponding edge (s,t) divided by the number of overall spanning trees $|\mathcal{T}|$.

One can see that by setting the weights ρ_{st} to one, the TRBP algorithm reduces to ordinary BP. In fact, all weights are equal to one if and only if the graph is a tree.

The TRBP method has a strong connection to a tree-relaxed linear program (LP) formulation of the MAP problem and one can provide an upper bound of the MAP probability after each iteration. This allows for estimating the quality of the estimated MAP configuration. One simply considers the gap between the estimated MAP probability (which is naturally an upper bound) and the lower bound, referred to as the *primal-dual gap*. If the primal-dual gap vanishes, the LP-relaxation is *tight* and one is guaranteed to have found the exact MAP. TRBP is, similar to loopy belief propagation, not guaranteed to converge. In practice, convergence can be enforced by introducing *message damping* with a damping factor $\alpha \in (0,1]$:

$$m_{t\to s}(x_s) = \left(\max_{x_t \in \mathcal{X}_t} \phi_{s,t}(x_s, x_t)^{\frac{1}{\rho_{st}}} \overline{m}_{t\to s}(x_t) \right)^{\alpha} m^{\text{old}}_{t\to s}(x_s)^{1-\alpha} \tag{3.27}$$

with *premessage*

$$\overline{m}_{t\to s}(x_t) \propto \phi_t(x_t) \frac{\prod\limits_{r \in \mathcal{N}_t \setminus \{s\}} m_{r\to t}(x_t)^{\rho_{st}}}{m_{s\to t}(x_t)^{1-\rho_{st}}} \tag{3.28}$$

In the last decade, there appeared a number of other dual BP approaches [115, 116, 63, 41, 45, 80, 47, 46]. The sequential tree-reweighted message passing algorithm proposed by Kolmogorov [63], for example, guarantees a monotonically decreasing upper bound. The method of [45] is another tree-reweighted message passing variant in which convergence is guaranteed. In practice, we observed that these methods require more iterations than TRBP in combination with a sufficiently large message damping. Other BP approaches such as [101] exploit the decomposability of the dual formulation to implement distributed variants. This allows to process very large graphical models where the full graph does not fit into the memory of a single computer but can be split into smaller subgraphs and distributed across several machines. An approach similar in spirit to the distributed BP method but with a more flexible formulation is provided in the following section.

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

Dual Decomposition

The family of inference methods based on dual decomposition (DD) [12] is discussed in the following. Instead of solving the MAP problem on the complete graph, one resorts to inference on smaller (independent) sub-problems, which only involves (overlapping) subsets of random variables. Each sub-problem (*slave*) works on its own copy of some *shared* random variables. A *master* forces *consensus* on the shared state configurations by iteratively balancing cost terms. The property of having independent sub-problems renders DD an ideal candidate for *parallelization* and *distributed computing*.

An intuitive analogy is provided via the resource allocation problem in economics. The resource allocation problem is the assignment of a limited amount of available resources to produce various goods with the goal to fit the needs of the society. The economics approach to solve this problem is via *pricing*. Global adjustment takes place via the free market price mechanism. Highly demanded resources become more expensive whereas underutilized resources get cheaper.

The methods proposed in this Section differ from the convex belief propagation methods mentioned in the previous Section in two ways. First, the sub-problems require only to solve MAP estimates instead of marginals. Second, the sub-problems are not restricted to tree structures. In fact, each sub-problem can have its own MAP inference method tailored to the specific problem structure. For example, consider joint image segmentation and pose estimation. The corresponding graphical model consists of two sets of random variables. One set of random variables represents the foreground/background segmentation labeling over image pixels and the other set encodes the pose of an object within the image. Each object pose random variable is connected with each pixel random variable. The image segmentation sub-problem is efficiently solved using graph cuts, whereas the pose estimation sub-problem is solved using belief propagation. In Sect. 5.3 this type of problem is described in more detail.

DD is a fairly old technique in the optimization community and somewhat surprisingly it has only recently experienced a revival in the computer vision and machine learning community. Its advantages are high flexibility, modularity, and promising convergence properties. Some recent computer vision problems solved via DD are multi-layer human pose estimation [31, 32], joint pose estimation and image segmentation [119] and higher-order graph matching [131].

In the following is provided a compact mathematical introduction to DD. The mathematical backbone of DD is Lagrangian relaxation [12, 65, 66]. The starting point of our derivation is MAP inference in a graphical model. For the sake of simplicity and to reduce notational clutter this introduction is restricted to pairwise MRFs, although an extension to higher-order models is straight forward.

$$\min_{\mathbf{x}\in\mathcal{X}} E(\mathbf{x}) = \sum_{s\in\mathcal{V}} \psi_s(x_s) + \sum_{(s,t)\in\mathcal{E}} \psi_{s,t}(x_s, x_t). \tag{3.29}$$

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

The decomposition methods do not directly work on the discrete state space $\mathcal{X}$ but on a reparametrization using *indicator variables*. Assume without loss of generality that the discrete state space is $\mathcal{X}_s = \{1, ..., L_s\}$ with L_s states per random variable X_s. The corresponding indicator variable vectors $\mathbf{y}_s = \{y_{sk}\}_{k=1}^{L_s}$, $\mathbf{y}_{st} = \{y_{stkl}\}_{k,l=1}^{L_s,L_t}$ for each vertex $s \in \mathcal{V}$ and for each edge $(s,t) \in \mathcal{E}$ are then defined as

$$y_{sk} = \begin{cases} 1 & \text{if } x_s = k \\ 0 & \text{otherwise,} \end{cases} \qquad y_{stkl} = \begin{cases} 1 & \text{if } x_s = k \wedge x_t = l \\ 0 & \text{otherwise.} \end{cases} \tag{3.30}$$

The unary and pairwise potential functions can then be reformulated as linear functions:

$$\psi_s(x_s) = \sum_{k=1}^{L_s} \theta_{sk} y_{sk} \qquad \psi_{s,t}(x_s,x_t) = \sum_{k=1}^{L_s} \sum_{l=1}^{L_t} \theta_{stkl} y_{stkl}. \tag{3.31}$$

with parameters θ. The resulting problem formulation then is:

$$\min_{\mathbf{y} \in \mathcal{M}(\mathcal{G},\mathcal{X})} E(\mathbf{y}) = \sum_{k=1}^{L_s} \theta_{sk} y_{sk} + \sum_{k=1}^{L_s} \sum_{l=1}^{L_t} \theta_{stkl} y_{stkl}. \tag{3.32}$$

where $\mathcal{M}(\mathcal{G},\mathcal{X})$ is the *marginal polytope*, that is the set of valid marginals $\{y_{sk}, y_{stkl} \in \mathbb{R} \mid \exists p \text{ such that } y_{sk} = \nu_s(X_s = k), y_{stkl} = \nu_{s,t}(X_s = k \wedge X_t = l)\}$. Note that the marginal polytope is the convex hull of valid (integer) configurations $\mathbf{y}$ and that minimizer of problem (3.32) are always integer [117].

Packing all elements y_{sk} and y_{stkl} into a single vector $\mathbf{y}$ and doing the same with the parameters, one can easily see that the equation in problem (3.32) is a scalar product. The simplified problem statement is thus

$$\min_{\mathbf{y} \in \mathcal{M}(\mathcal{G},\mathcal{X})} \langle \theta, \mathbf{y} \rangle. \tag{3.33}$$

The structure of the marginal polytope $\mathcal{M}(\mathcal{G}, \mathcal{X})$ can be quite complex. Problem (3.33) is an integer program which in general is NP-hard [117]. However, there exist a number of structures for which efficient algorithms for solving (3.33) exist (e.g. tree-structured graphs or sub-modular objective functions).

The approach for decomposition methods is to decompose the problem (3.33) into a sum of tractable sub-terms:

$$\min_{\mathbf{y} \in \mathcal{M}(\mathcal{G},\mathcal{X})} E(\mathbf{y}) = \sum_{\tau \in \mathcal{T}} E_\tau(\mathbf{y}) = \sum_{\tau \in \mathcal{T}} \langle \theta_\tau, \mathbf{y} \rangle \tag{3.34}$$

The sub-terms can again be defined over (sub-)graphs $\mathcal{G}_\tau = (\mathcal{V}_\tau, \mathcal{E}_\tau)$:

$$E_\tau(\mathbf{y}_\tau) = \langle \theta_\tau, \mathbf{y}_\tau \rangle \tag{3.35}$$

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

subject to the marginalization constraints $\mathcal{M}(\mathcal{G}_\tau, \mathcal{X})$. Obviously, the parameters θ_τ have to be chosen such that the original distribution is not altered, that is: $\theta = \sum_{\tau \in \mathcal{T}} \theta_\tau$. The main idea of dual decomposition now is to introduce a copy of the configuration vector for each sub-term and to add an equality constraint which enforces consistency:

$$\min_{\mathbf{y}, \{\mathbf{y}_\tau\}_{\tau\in\mathcal{T}}} \quad \sum_{\tau\in\mathcal{T}} E_\tau(\mathbf{y}_\tau) \tag{3.36}$$

$$\text{s.t.} \quad \mathbf{y}_\tau \in \mathcal{M}(\mathcal{G}_\tau, \mathcal{X}) \ \forall \tau \in \mathcal{T} \tag{3.37}$$

$$\mathbf{y}_\tau = \mathbf{y} \ \forall \tau \in \mathcal{T}. \tag{3.38}$$

Decomposition into independent subproblems is achieved by relaxing the hard-constraint (3.38) using Lagrangian relaxation. The Lagrangian dual is equivalent to an LP relaxation where the marginal polytope $\mathcal{M}(\mathcal{G}, \mathcal{X})$ is replaced by a looser (much simpler) constraint set $\mathcal{L}(\mathcal{G}, \mathcal{X}) \supseteq \mathcal{M}(\mathcal{G}, \mathcal{X})$, called the *local polytope* [116].

The resulting subproblems can then be solved independently from each other. The decomposition of problem (3.33) has to be done such that the resulting subproblems have pleasant properties which allow for efficient inference. For example, tree-structured subproblems can be efficiently solved with the max-product BP algorithm in Sec. 3.3.1.

MAP-DD

In the DD of [65, 66], Lagrangian relaxation is applied on the consistency constraint (3.38), leading to the following problem formulation:

$$\max_{\lambda\in\Lambda} \min_{\mathbf{z}, \{\mathbf{y}_\tau\}_\tau} \mathcal{L}(\{\mathbf{y}_\tau\}_\tau, \mathbf{z}, \lambda) \tag{3.39}$$

where $\lambda \in \Lambda$ are Lagrange multipliers restricted to a convex *feasible set* $\Lambda = \{\lambda_\tau \mid \sum_{\tau\in\mathcal{T}} \lambda_\tau = \mathbf{0}\}$. The variable $\mathbf{y}$ is replaced by $\mathbf{z}$ in order to emphasize that the optimal solution $\mathbf{z}^*$ of Eq. (3.39) is not necessarily an optimizer (or even a feasible solution) to the original MAP problem. This is especially the case when $\mathbf{z}^*$ is not integer. The Lagrangian $\mathcal{L}(\mathbf{y}, \mathbf{z}, \lambda)$ is as follows:

$$\mathcal{L}(\{\mathbf{y}_\tau\}_\tau, \mathbf{z}, \lambda) = \sum_{\tau\in\mathcal{T}} \left[\langle \theta_\tau, \mathbf{y}_\tau \rangle + \lambda_\tau^\mathsf{T}(\mathbf{y}_\tau - \mathbf{z})\right]. \tag{3.40}$$

Rearranging equation (3.40) leads to

$$\mathcal{L}(\{\mathbf{y}_\tau\}_\tau, \mathbf{z}, \lambda) = \sum_{\tau\in\mathcal{T}} \langle \theta_\tau + \lambda_\tau, \mathbf{y}_\tau \rangle - (\sum_{\tau\in\mathcal{T}} \lambda_\tau)^\mathsf{T}\mathbf{z}. \tag{3.41}$$

Since the minimization over $\mathbf{z}$ is unconstrained, it is easy to see that the inner (minimization) problem of (3.39) is only bounded when the right-most term of

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

Eq. (3.41), $(\sum_{\tau\in\mathcal{T}}\lambda_\tau)^\intercal\mathbf{z}$, is equal to zero for all $\mathbf{z}$. This is exactly the case when $\sum_{\tau\in\mathcal{T}}\lambda_\tau = \mathbf{0}$, and thus $\lambda\in\Lambda$:

$$\min_{\mathbf{z},\{\mathbf{y}_\tau\}_\tau}\mathcal{L}(\{\mathbf{y}_\tau\}_\tau,\mathbf{z},\lambda) = \begin{cases} g(\lambda) & \text{if } \lambda\in\Lambda \\ -\infty & \text{otherwise} \end{cases} \tag{3.42}$$

with the dual function $g(\lambda) = \min\limits_{\{\mathbf{y}_\tau\}_\tau}\sum_{\tau\in\mathcal{T}}\langle\theta_\tau+\lambda_\tau,\mathbf{y}_\tau\rangle$. That is, a λ which maximizes $g(\lambda)$ must lie in the subspace Λ. The minimization of the dual g is now completely independent of $\mathbf{z}$. This is exactly the property which enables splitting into independent subproblems:

$$g(\lambda) = \sum_{\tau\in\mathcal{T}} g_\tau(\lambda_\tau) \qquad\qquad g_\tau(\lambda_\tau) = \min_{\mathbf{y}_\tau\in\mathcal{M}(\mathcal{G}_\tau,\mathcal{X})}\langle\theta_\tau+\lambda_\tau,\mathbf{y}_\tau\rangle. \tag{3.43}$$

The *weak duality theorem* guarantees that $g(\lambda)\le E(\mathbf{y}) = \langle\theta,\mathbf{y}\rangle$ for all λ and for all $\mathbf{y}$. That is, $g(\lambda)$ lower bounds $E(\mathbf{y})$. The aim is to reduce the *gap* between the lower bound (the dual energy) $g(\lambda)$ and the current MAP energy $E(\mathbf{y})$ (the primal energy) as much as possible.

The dual problem of the primal problem (3.32) then is:

$$\max_{\lambda\in\Lambda} g(\lambda) = \sum_{\tau\in\mathcal{T}} g_\tau(\lambda_\tau). \tag{3.44}$$

It is easy to verify that the negative dual $-g(\lambda)$ is convex. Since $g(\lambda)$ is not continuously differentiable, usual gradient ascent algorithms are not applicable. Additionally, the constraint $\lambda\in\Lambda$ has to be fulfilled.

These requirements lead to the following *projected subgradient method*:

$$\lambda_\tau^{n+1} = [\lambda_\tau^n + \alpha_n\nabla g_\tau(\lambda_\tau^n)]_\Lambda \tag{3.45}$$

where $-\nabla g_\tau(\lambda_\tau)$ is a *subgradient* of $-g_\tau(\,\cdot\,)$ at point λ_τ and n is the current DD iteration. The positive constant α_n denotes the step length for gradient ascend and $[\,\cdot\,]_\Lambda$ describes the projection onto the feasible set Λ.

A subgradient $\nabla f(x_0)$ of a convex function $f(x)$ at point x_0 is the slope of a linear function (the subtangent) such that the following condition holds:

$$f(x) \ge f(x_0) + \nabla f(x_0)\cdot(x-x_0) \qquad\qquad \forall x. \tag{3.46}$$

A subgradient is not necessarily unique. The set of subgradients at point x_0 is called the *subdifferential at point* x_0, denoted by $\nabla f(x_0)\in\partial f(x_0)$.

Somewhat surprisingly, it turns out that $\nabla g_\tau(\lambda_\tau) = \mathbf{y}_\tau^*$ is a subgradient of $-g_\tau(\lambda_\tau)$, where $\mathbf{y}_\tau^*$ is the minimizer of the slave problem in Eq. (3.43) [65]. The projection operator $[\,\cdot\,]_\Lambda$ reduces to subtracting the average vector $\bar\lambda = \frac{1}{|\mathcal{T}|}\sum_\tau\lambda_\tau$ from each λ_τ. The final projected subgradient update can be summarized as follows:

$$\lambda_\tau^{n+1} = [\lambda_\tau^n + \alpha_n\mathbf{y}_\tau^*]_\Lambda = \lambda_\tau^n + \alpha_n(\mathbf{y}_\tau^* - \frac{1}{|\mathcal{T}|}\sum_{\tau'\in\mathcal{T}}\mathbf{y}_{\tau'}^*). \tag{3.47}$$

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

The choice of the step size α_n crucially influences the convergence behavior of the DD algorithm. One can show that by using step size sequences which satisfy the following conditions, the sub-gradient algorithm converges to the optimal solution of the relaxed problem in Eq. (3.44) [65].

$$\alpha_n \geq 0 \qquad \lim_{n\to\infty} \alpha_n = 0 \qquad \sum_{n=0}^{\infty} \alpha_n = \infty \tag{3.48}$$

It is guaranteed that after $N = \mathcal{O}(1/\epsilon^2)$ iterations, ϵ-accuracy is reached with $g(\lambda^N) - g(\lambda^*) \leq \epsilon$, where λ^* is the optimal solution [12]. There exist various choices on selecting the step size [65, 66]. A time complexity of $\mathcal{O}(1/\epsilon^2)$ is impractically slow. An accelerated DD method has been proposed in [55] based on smoothing the Lagrangian relaxation, achieving a much better time complexity of $\mathcal{O}(1/\epsilon)$. The (ordinary) DD algorithm for solving the relaxed MAP problem is summarized in Alg. 3.

Algorithm 3 MAP Dual Decomposition Algorithm

```
Input: Problem decomposition $\{\theta_\tau\}_{\tau\in\mathcal{T}}$, step sizes $\{\alpha_n\}_n$
Ensure: MAP estimate $\mathbf{y}$
Initialize $\lambda_\tau^1 = \mathbf{0}$
for $n = 1$ to $N$ do
    for Subproblems $\tau \in \mathcal{T}$ do
        Solve MAP $\mathbf{y}_\tau^{n+1} = \arg\min_{\mathbf{y}_\tau \in \mathcal{M}(\mathcal{G}_\tau, \mathcal{X})} \langle \theta_\tau + \lambda_\tau^n, \mathbf{y}_\tau \rangle$
    end for
    Compute consensus $\mathbf{z}^{n+1} = \frac{1}{|\mathcal{T}|} \sum_{\tau\in\mathcal{T}} \mathbf{y}_\tau^{n+1}$
    Update Lagrange multipliers $\lambda_\tau^{n+1} = \lambda_\tau^n + \alpha_n(\mathbf{y}_\tau^{n+1} - \mathbf{z}^{n+1})$
end for
Decode primal estimate $\mathbf{y}$ from partial solutions $\mathbf{y}_\tau$
```

The methods introduced so far handle MAP inference in graphical models over discrete random variables. The goal in this thesis is to perform inference over *continuous* random variables. A key technique to achieve this goal is via Monte-Carlo simulation which is the topic of the following two sections.

3.4 Markov Chain Monte-Carlo Methods

In the following, first the idea of Monte-Carlo simulation is briefly introduced, followed by Markov chain Monte-Carlo (MCMC) simulation (see [122, 3] for an in-depth introduction). This introduction focuses on the most fundamental MCMC methods: the Gibbs sampler and the Metropolis-Hastings sampler. After this, the slice sampler [88] is introduced which provides the basis for our proposed MAP inference approach in Chapt. 4. See [3] for a more detailed introduction on MCMC methods for machine learning.

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

Monte-Carlo simulation is a sampling based approach to approximately calculate (or simulate) certain properties, such as, e.g., the expectation

$$E_\mu[g(X)] = \int_{\mathcal{X}} g(x)\mu(x)dx \tag{3.49}$$

of a real-valued function g. The idea is to approximate the target density μ by the *empirical point-mass function*

$$\mu_N(x) = \frac{1}{N}\sum_{i=1}^{N} \delta_{x^{(i)}}(x) \tag{3.50}$$

using a finite set of independently and identically distributed (i.i.d) drawn samples $\{x^{(i)}\}_{i=1}^N$ from the target density $\mu(x)$. Here, $\delta_{x^{(i)}}(x)$ denotes the Dirac impulse located at $x^{(i)}$. The expectation $E_{\mu_N}[g(X)]$ converges almost surely (a.s.) to $E_\mu[g(X)]$ for $N \to \infty$:

$$E_{\mu_N}[g(X)] = \frac{1}{N}\sum_{i=1}^{N} g(x^{(i)}) \xrightarrow[\text{a.s.}]{N\to\infty} \int_{\mathcal{X}} g(x)\mu(x)dx = E_\mu[g(X)], \tag{3.51}$$

according to the strong law of large numbers.

An important advantage of Monte-Carlo simulation towards deterministic integration is that the discretization points (i.e. the samples) concentrate in regions of high probability. Thus, one can also use these samples for approximate MAP inference:

$$\arg\max_{x^{(i)}} \mu(x). \tag{3.52}$$

This behavior will be of fundamental use to construct efficient stochastic MAP inference algorithms as presented in chapter 4.

Direct or exact sampling from the target distribution $\mu(x)$ is in general not feasible. In many cases, it is much easier to sample from a conditional density $p(x^{(i)} \mid x^{(i-1)})$. This leads to the idea of MCMC sampling.

A Markov chain is a special case of a stochastic process. More precisely, it is a sequence of random variables $x^{\langle 0\rangle}, x^{\langle 1\rangle}, ... \in \mathcal{X}$ where the conditional distribution of $x^{\langle m\rangle}$ given $x^{\langle 0\rangle}, ..., x^{\langle m-1\rangle}$ depends on the last element $x^{\langle m-1\rangle}$ only, i.e.

$$p(x^{\langle m\rangle} \mid x^{\langle 0\rangle}, ..., x^{\langle m-1\rangle}) = p(x^{\langle m\rangle} \mid x^{\langle m-1\rangle}). \tag{3.53}$$

The first element $x^{\langle 0\rangle}$ is has an arbitrary initial probability (density) $\nu_0(x^{\langle 0\rangle})$. The second element depends on the first element by the conditional distribution $T_1(x^{\langle 0\rangle}, x^{\langle 1\rangle}) = p(x^{\langle 1\rangle} \mid x^{\langle 0\rangle})$, the *transition probability density* or the *transition kernel.* The joint probability $p(x^{\langle 0\rangle}, ..., x^{\langle m\rangle})$ is obtained using the chain rule:

$$p(x^{\langle 0\rangle}, ..., x^{\langle m\rangle}) = \nu_0(x^{\langle 0\rangle})T_1(x^{\langle 0\rangle}, x^{\langle 1\rangle}) \cdot ... \cdot T_{m-1}(x^{\langle m-1\rangle}, x^{\langle m\rangle}). \tag{3.54}$$

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

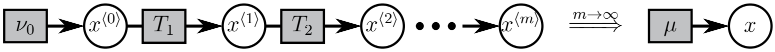


Figure 3.4: Graphical model for Markov chain Monte-Carlo sampling.

The marginal probability of the second element is:

$$\nu_2(x^{\langle 1\rangle}) = \int_{\mathcal{X}} \nu_0(x^{\langle 0\rangle})T_1(x^{\langle 1\rangle}, x^{\langle 0\rangle})dx^{\langle 0\rangle}. \tag{3.55}$$

Likewise, the marginal probability of the m-th element is:

$$\nu_m(x^{\langle m\rangle}) = \int_{\mathcal{X}\times\cdots\times\mathcal{X}} \nu_0(x^{\langle 0\rangle})T_1(x^{\langle 0\rangle}, x^{\langle 1\rangle})\cdots T_{m-1}(x^{\langle m-1\rangle}, x^{\langle m\rangle})dx^{\langle 0\rangle}\cdots dx^{\langle m-1\rangle}. \tag{3.56}$$

A Markov chain is called *homogeneous* if the transition kernel T_m does not depend on m, i.e. $T = T_m\ \forall m$. A Markov chain is called *irreducible* if when starting from any state it reaches eventually each other state with positive probability (density). A Markov chain is called *aperiodic* if it can not get trapped in cycles. If a Markov chain is irreducible and aperiodic, then the Markov chain converges to a unique stationary distribution.

Our goal is to construct a transition kernel T such that the marginal distribution $\nu_m(x^{\langle m\rangle})$ asymptotically converges to the target distribution $\mu(x)$ for $m \to \infty$. The convergence should be independent of the starting point ν_0, the (initial) distribution of $x^{\langle 0\rangle}$. Figure 3.4 summarizes the MCMC approach graphically.

The question is how to construct transition kernels T such that the Markov chain converges to the desired target function μ? An approach is given using the notion of *reversibility.*

A Markov chain is called *reversible* with respect to the distribution $\mu(x^{\langle m-1\rangle})$ if the *detailed balance* (d.b.) condition is fulfilled:

$$\mu(x^{\langle m\rangle})T(x^{\langle m\rangle}, x^{\langle m-1\rangle}) = \mu(x^{\langle m-1\rangle})T(x^{\langle m-1\rangle}, x^{\langle m\rangle}). \tag{3.57}$$

The detailed balance condition is sufficient for convergence to the target distribution $\mu(x)$. This can be seen by marginalization over $x^{\langle m-1\rangle}$:

$$\int\limits_{\mathcal{X}} \mu(x^{\langle m-1\rangle})T(x^{\langle m-1\rangle}, x^{\langle m\rangle})dx^{\langle m-1\rangle} \overset{\text{d.b.}}{=} \int\limits_{\mathcal{X}} \mu(x^{\langle m\rangle})T(x^{\langle m\rangle}, x^{\langle m-1\rangle})dx^{\langle m-1\rangle} = \mu(x^{\langle m\rangle}). \tag{3.58}$$

The right-most equality is due to the property $\int_{\mathcal{X}} T(x,y)dy = 1$ for conditional probabilities. Omitting the variables results in the simplified formula $\mu T = \mu$. That is, the target distribution μ is a fix-point of the Markov chain. Note that the Markov chain need not be homogeneous in order to provide convergence. It is sufficient that detailed balance holds for each T_m.

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

3.4.1 Metropolis-Hastings Sampler

This method has been widely used for approximate inference, especially the particle belief propagation (PBP) approach in Sect. 3.5.

An Metropolis-Hastings (MH) step consists of two parts. The first part is sampling from a *proposal* distribution $q(x^{\langle m \rangle} \mid x^{\langle m-1 \rangle})$. The generated sample x^* is a *candidate*. The proposal distribution only needs to fulfill some fairly general conditions and are generally designed to be *easy* to sample from. A very popular family of proposal distributions is the Gaussian distribution. The second step is an *acceptance test*. The candidate sample x^* is accepted with probability

$$\min\left\{1, \frac{\mu(x^*)q(x^{\langle m-1 \rangle} \mid x^*)}{\mu(x^{\langle m-1 \rangle})q(x^* \mid x^{\langle m-1 \rangle})}\right\}. \tag{3.59}$$

One can show that the induced transition kernel satisfies the detailed balance condition for the target distribution μ [3]. Convergence of MH is ensured if aperiodicity and irreducibility is provided. Aperiodicity is ensured due to the rejection step (a candidate can be rejected at any time, having the effect of "breaking" cycles). Irreducibility is provided when the support of the proposal distribution q includes the support of μ, i.e. $\{x \,:\, \mu(x) > 0\} \subset \{x \,:\, q(x \mid y) > 0\}$.

Figures 3.5–3.7 show an example MH simulation for the Gaussian mixture distribution $p(x) = 0.6 \cdot \mathcal{N}(x; -1.5, 0.5) + 0.4 \cdot \mathcal{N}(x; 2.5, 2.0)$. These figures show how important it is to select an appropriate proposal distribution. If the proposal distribution is too narrow (cf. Fig. 3.6), the MCMC chains move too slowly and thus the sampling space is not discovered well. If the proposal distribution is too broad (cf. Fig. 3.7), the samples get rejected too often and the MCMC chains "get stuck".

A huge variety of other sampling methods have been proposed over the years. Listing these approaches is out of scope of this thesis. A very promising approach, slice sampling, is presented in Sect. 3.4.3. The slice sampler is much more robust against parameter choice than MH. In some special cases, this method can be made parameter-free. But before explaining slice sampling we need to introduce another very important sampler, the Gibbs sampler, in the following section.

3.4.2 Gibbs Sampler

Let us for now consider sampling from a joint distribution with random variables $\mathbf{X} = (X_1, ..., X_N)^\mathsf{T}$. In Gibbs sampling, a single element (or a subset of elements) s is picked from the current configuration vector $\mathbf{x}^{\langle m-1 \rangle}$ in turn and sampling is done solely on this selected subset while all other elements are kept fixed. An advantage of this method is that sampling from the conditional distributions $p(x_s \mid x_{-s})$ can often be implemented very efficiently.

The Gibbs sampler can be interpreted as a special case of the MH sampler with

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

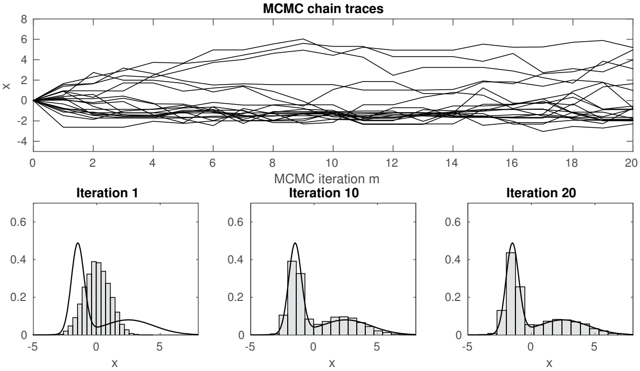


Figure 3.5: MH-MCMC simulations of a toy Gaussian mixture distribution repeated 10 000 times. Top: The first twenty Markov chains of the MH simulation using a Gaussian proposal distribution with standard deviation of 1.0. Bottom: Histograms and target distribution at different simulation steps. The histogram fits the target distribution well after twenty MCMC iterations.

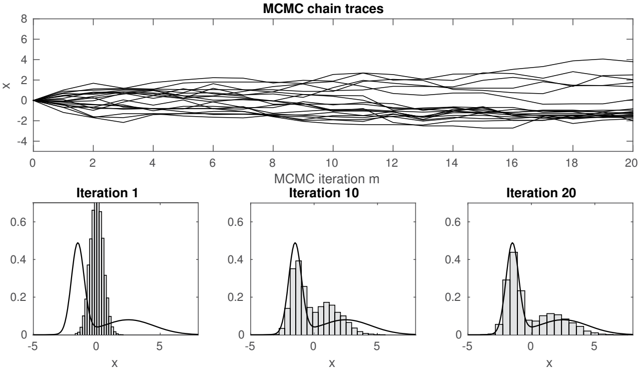


Figure 3.6: MH-MCMC simulations using a Gaussian proposal distribution with standard deviation of 0.5. The MCMC chains move very slowly, degrading the sample quality.

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

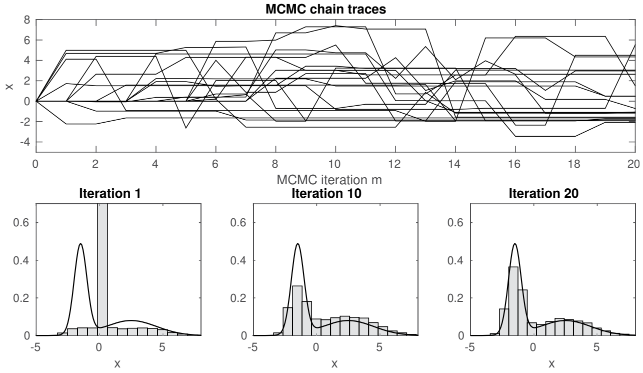


Figure 3.7: MH-MCMC simulations using a Gaussian proposal distribution with standard deviation of 10.0. The MCMC chains often get stuck due to a high sample rejection rate. This again degrades the sample quality.

proposal distribution

$$q_s(\mathbf{x}^* \mid \mathbf{x}^{\langle m-1\rangle}) = \begin{cases} p(x_s^* \mid \mathbf{x}_{-s}^{\langle m-1\rangle}) & \text{if } \mathbf{x}_{-s}^* = \mathbf{x}_{-s}^{\langle m-1\rangle} \\ 0 & \text{else.} \end{cases} \tag{3.60}$$

The vector $\mathbf{x}_{-s}$ denotes the subvector $(x_1, ..., x_{s-1}, x_{s+1}, ..., x_N)^\mathsf{T}$ (i.e. vector $\mathbf{x}$ without entry x_s).

It is easy to see that the acceptance probability is always 1, that is there is no rejection of sample candidates. If one can sample directly from the conditional distributions $p(x_s^* \mid x_{-s}^{\langle m-1\rangle})$ for all vertices s, then Gibbs sampling is a very efficient (and often easy to implement) method.

The main disadvantage of this method is that convergence of the Markov chain can be very slow if the random variables are highly correlated. One way to counter this issue is to sample from a block of variables, leading to *block Gibbs sampling*. Here, s is redefined as a set of vertices $s \subset \{1, ..., N\}$. Formula (3.60) applies analogously.

3.4.3 Slice Sampler

In this section the concept of slice sampling [88, 3] is introduced. This method is an instance of so-called auxiliary variable samplers. Instead of sampling from the target distribution $p(x)$ with state space $\mathcal{X}$, one instead samples from a higher-dimensional distribution $q(x, u)$ with state space $\mathcal{X} \times \mathcal{U}$.

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

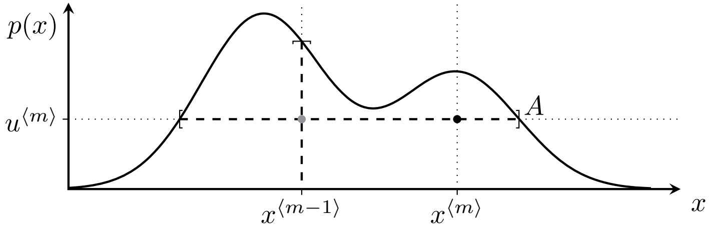


Figure 3.8: Slice Sampling [88, 3]. Given a previous sample point $x^{\langle m-1 \rangle}$, first an auxiliary variable $u^{\langle m \rangle}$ is sampled uniformly from the interval $[0, q(x^{\langle m-1 \rangle})]$. Afterwards, the new sample $x^{\langle m \rangle}$ is uniformly sampled from the region for which $q(x^{\langle m \rangle}) > u^{\langle m \rangle}$, defining the *slice* A.

The joint distribution is as follows

$$q(x, u) = \begin{cases} \frac{1}{Z} & \text{if } 0 < u < q(x) \\ 0 & \text{otherwise} \end{cases} \tag{3.61}$$

for any $q(x) \propto p(x)$ and normalization constant $Z = \int q(x)dx$. Intuitively, this approach can be interpreted as uniformly sampling from the volume under the plot of $q(x)$, i.e. $V = \{x \in \mathcal{X}, u \in \mathcal{U} \ : \ 0 < u < p(x)\} \subset \mathcal{X} \times \mathcal{U}$.

Direct sampling from this density function is in general not feasible. Therefore, one resorts to Gibbs sampling, alternately sampling from the conditional distributions $q(u \mid x)$ and $q(x \mid u)$. The conditional distributions are particularly simple:

$$u^{\langle m \rangle} \sim q(u \mid x^{\langle m-1 \rangle}) = \mathcal{U}\left[0, q(x^{\langle m-1 \rangle})\right] \tag{3.62}$$

$$x^{\langle m \rangle} \sim q(x \mid u^{\langle m \rangle}) = \mathcal{U}(A_q(u^{\langle m \rangle})), \tag{3.63}$$

where $\mathcal{U}\left[l, r\right]$ is the uniform distribution over an interval with lower bounds l and upper bounds r, $\mathcal{U}(A)$ is the uniform distribution over a region A, and the region $A_q(u^{\langle m \rangle})$ is defined as

$$A_q(u^{\langle m \rangle}) = \{x; \ q(x) \geq u^{\langle m \rangle}\}. \tag{3.64}$$

The first step can be interpreted as selecting a level set (i.e. a *slice*) u at which to cut through the plot of the graph, as shown in Fig. 3.8. The second step is uniformly sampling from the region for which $q(x) > u$.

This method is easy to implement and very efficient given that the interval A is easy to calculate. Unfortunately, this is rarely the case.

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

An extension of this approach is the *product slice sampler* [83, 26]. Assume that $q(x)$ can be decomposed in L functions $q_l(x)$ such that

$$q(x) \propto \prod_{l=1}^{L} q_l(x) \tag{3.65}$$

and $q_l(x)$ is *easy* in terms of $\{x;\ q_l(x) \geq u\}$ can be computed efficiently. Then a sample from $q(x)$ can be drawn by introducing L auxiliary variables $u_1, \ldots, u_L$:

$$u^{\langle m \rangle}{}_1 \sim q(u_1 \mid x^{\langle m-1 \rangle}) = \mathfrak{U}\left[0, q_1(x^{\langle m-1 \rangle})\right] \tag{3.66a}$$

$$\vdots$$

$$u^{\langle m \rangle}{}_L \sim q(u_L \mid x^{\langle m-1 \rangle}) = \mathfrak{U}\left[0, q_L(x^{\langle m-1 \rangle})\right] \tag{3.66b}$$

$$x^{\langle m \rangle} \sim q(x \mid u^{\langle m \rangle}{}_1, \ldots, u^{\langle m \rangle}{}_L) = \mathfrak{U}(A_q(u^{\langle m \rangle})), \tag{3.66c}$$

where $A_q(u^{\langle m \rangle}) = \{x\ ;\ q_l(x) \geq u_l^{\langle m \rangle}, l = 1, \ldots, L\}$ [3].

The main difficulty lies in determining the slice region $A_q(u^{\langle m \rangle})$. A simple observation is that this region decomposes over an intersection of sub-regions $A_{q_l}(u_l^{\langle m \rangle}) = \{x\ ;\ q_l(x) \geq u_l^{\langle m \rangle}\}$ for each factor $q_l(x)$:

$$A_q(u^{\langle m \rangle}) = \bigcap_{l=1}^{L} A_{q_l}(u_l^{\langle m \rangle}). \tag{3.67}$$

Assuming that the sub-regions $A_{q_l}(u_l^{\langle m \rangle})$ can be computed (or at least approximated) efficiently, the implementation of product slice-sampling is straight forward. Some important function families for which this is the case are summarized in Sect. 4.1.2.

There have been proposed two methods by Neal [88] for approximating the slices in case that a direct computation of $A_{q_l}(u_l^{\langle m \rangle})$ is infeasible: the stepping-out method and the doubling method. These two approaches work by first finding an approximate interval $A_{q_l}(u_l^{\langle m \rangle}) \approx [L,R]$ with left and right interval bounds L and R, respectively, and then performing rejection sampling on the approximated interval in combination with an interval shrinking scheme in order to reduce the rejection rate. Note that care must be taken in the choice of the interval expansion and shrinking steps to ensure a valid MCMC scheme which converges to the target distribution $p(x)$. The resulting constraint is that the interval $[L, R]$ must be able to be constructed from point $x^{\langle m-1 \rangle}$ as likely as from point $x^{\langle m \rangle}$, i.e.

$$\begin{aligned} x^{\langle m \rangle} \in \{x\ | x \in A_{q_l}(u_l^{\langle m \rangle}) \cap [L,R] \text{ and} \\ P(\text{select } [L, R] \mid \text{at state } x) = P(\text{select } [L, R] \mid \text{at state } x^{\langle m-1 \rangle})\}. \end{aligned} \tag{3.68}$$

Two approaches for finding the initial interval bounds are proposed: “stepping out” and “doubling”. In “stepping out”, an initially drawn interval of fixed width

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

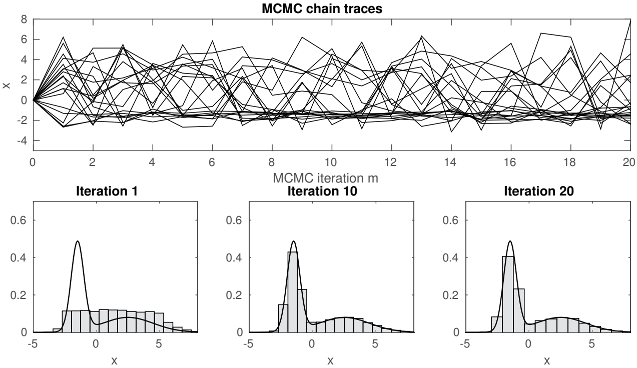


Figure 3.9: Slice sampling MCMC simulation of the same distribution as shown in Figs. 3.5–3.7. The MCMC chains jump more quickly than the MH chains (good chain mixing). After ten MCMC iterations, the target distribution is approximated better than with the MH approach.

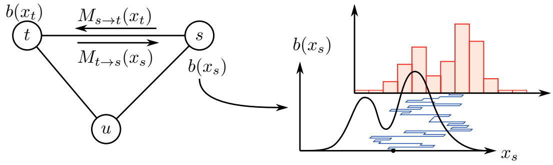
around $x^{\langle m-1\rangle}$ and iteratively widened by a constant step length until the outer bounds L and R are outside of $A_{q_l}(u_l^{\langle m\rangle})$. This can be tested by simple point-wise evaluation of $q_l(\,\cdot\,)$. The "doubling" procedure doubles in each iteration the width of the interval by expanding with equal probability either the left bound L to the left or the right bound R to the right until both bounds are outside of $A_{q_l}(u_l^{\langle m\rangle})$.

The shrinking operation is as follows: A new sample candidate x^* is generated from the interval $[L,R]$ and tested if $x^* \in A_{q_l}(u_l^{\langle m\rangle})$ using point-wise evaluation. If not, then the interval $[L,R]$ is successively shrunk and a new sample candidate is drawn. This shrinking operation continues until a valid sample is found. For the doubling procedure it may be necessary to further check for condition (3.68).

For multivariate sampling, i.e. $\mathbf{x} = (x_1, ..., x_N)^\mathsf{T}$, one straightforward method is to apply Gibbs sampling as in Sect. 3.4.2 [88]. The conditional distributions can then be sampled using single-variable slice sampling [26].

Figure 3.9 depicts the MCMC chain simulations from the example introduced in Sect. 3.4.1 but using slice sampling with the step-out procedure instead of MH. Slice sampling leads to much better chain mixing and hence reaches a good approximation of the target distribution earlier than MH (after the 10th iteration).

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.



Graphical Model (exemplary)

MCMC sampling

Figure 3.10: Particle Belief Propagation framework. Left: Message passing mechanism. Right: MCMC particle sampling of the belief $b(x_s)$ with an exemplary MCMC sampling chain of one particle (blue) and its corresponding histogram (red).

3.5 Max-Product Particle Belief Propagation

For arbitrary potential functions $\phi_s(x_s)$ and $\phi_{s,t}(x_s, x_t)$ over continuous variables, the max-operators in Eqs. (3.20) and (3.22) for the message passing and decoding operations are in general intractable. In some special cases, these operations can be solved exactly and computationally efficiently. This is, e.g., the case if all potential functions are Gaussian, thus leading to the special case of Gaussian Markov random fields [97] and especially to Gaussian BP [15].

In the following the max-product particle BP algorithm [67, 14] is summarized. One method for approximating the max-operator over the continuous state space $\mathcal{X}_t$ in the message-passing rule in Eq. (3.20) is to restrict the max-operation over a discrete and finite set of particles $\mathcal{P}_t = \{x_t^{(1)}, \dots, x_t^{(p)}\}$, where p is the number of particles per node. The modified max-product rule is:

$$m_{t\to s}(x_s) \approx \hat{m}_{t\to s}(x_s) = \max_{x_t \in \mathcal{P}_t} [\phi_t(x_t)\phi_{s,t}(x_s, x_t) \prod_{r \in \mathcal{N}_t \setminus \{s\}} \hat{m}_{r\to t}(x_t)]. \tag{3.69}$$

In fact, since $\mathcal{P}_t \subset \mathcal{X}_t$, it is $\hat{m}_{t\to s}(x_s) \leq m_{t\to s}(x_s)$ and thus the true max-marginal $\mu_s(x_s)$ is always underestimated:

$$\hat{\mu}_s(x_s) = \phi_s(x_s) \prod_{t \in \mathcal{N}_s} \hat{m}_{t\to s}(x_s) \leq \mu_s(x_s). \tag{3.70}$$

Nevertheless, approximate MAP estimates can be decoded from:

$$\hat{x}_s^* \in \arg\max_{x_s \in \mathcal{P}_s} \hat{\mu}_s(x_s). \tag{3.71}$$

Note that the belief $\mu_s(x_s)$ and the messages $\hat{m}_{t\to s}(x_s)$ can be calculated for all continuous values $x_s \in \mathcal{X}_s$ rather than only on the particle set $\mathcal{P}_s$. On the other hand, the messages from node s to node t are approximated only using the particles x_t from the particle set $\mathcal{P}_t$ of node t.

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

Algorithm 4 Particle belief propagation [67, 14]

```
Input: Initial set of particles: $\{x_s^{(i)}\}_{i=1,\dots,p}$, proposal distribution $q$
Initialize the messages $m^0_{t\to s}(x_s) = 1$ and beliefs $\mu^0_s(x_s^{(i)}) = 1$ for all $s$,$t$
for BP iteration $n = 1$ to $N$ do
    for each node $s$ and each particle $i = 1, \dots, p$ do
        Initialize sampling chain $x_s^{(i)\langle 0\rangle} \leftarrow x_s^{(i)}$
        for MCMC iteration $m = 1, \dots, M$ do
            Sample $\bar{x}_s^{(i)\langle m\rangle} \sim q_\sigma(x \mid x_s^{(i)\langle m-1\rangle})$
            Calc. belief $\mu^n_s(\bar{x}_s^{(i)\langle m\rangle})$ from Eqs. (3.70), (3.69)
            Sample $u \sim \mathcal{U}\,[0,1]$
            if $u \leq \min\left\{1, {\mu^n_s(x_s^{(i)\langle m\rangle})}/{\mu^n_s(\bar{x}_s^{(i)\langle m\rangle})}\right\}$ then
                Accept: $x_s^{(i)\langle m\rangle} \leftarrow \bar{x}_s^{(i)\langle m\rangle}$
            end if
        end for
        $x_s^{(i)} \leftarrow x_s^{(i)\langle M\rangle}$
    end for
    Normalize messages and beliefs
end for
```

A major issue in particle-based inference methods is how and from which distribution to sample the particles. Koller [62] suggests that the true marginal $\nu_s(x_s)$ would be the best choice. This statement is mathematically founded by [50]. Since this quantity is not available, one resorts to sampling from the most recent max-marginal estimate $\mu^n_s(x_s)$ [67]. Messages and beliefs are calculated iteratively for $n = 1, \dots, N$ iterations. New particles are sampled in each BP iteration using the most recent belief estimate.

The main issue in PBP is then how to sample from the (arbitrarily shaped) belief density function. For this task a short MH-MCMC chain simulation is typically used. This method requires a proposal distribution q where new particles can be easily sampled from. Typically a Gaussian function $q(x \mid y) = q_\sigma(x \mid y)$ with a predefined standard deviation σ is used.

Figure 3.10 shows a schematic overview of the PBP framework. Algorithm 4 summarizes the Metropolis-Hastings based max-product belief propagation algorithm (MH-PBP).

The proposal function q needs to be carefully adjusted to the true belief distribution. This introduces a dependency on prior knowledge about how the labels are distributed in the label space. In Sec. 4.1, a modification of algorithm 4 is proposed which replaces the MH sampler by an appropriate slice sampler which exploits the message-passing structure for efficient sampling.

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

Chapter

4

Stochastic Inference in Probabilistic Graphical Models

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

In recent years, stochastic inference methods have been applied for solving maximum a posteriori (MAP) inference problems in high-dimensional, continuous or discrete-continuous probabilistic networks [62, 112, 92, 14, 91]. These methods face two major difficulties. To explain them, consider the following naive approach: First, randomly generate sample proposals from the set of feasible solutions. Then, evaluate the joint distribution at these configurations and return the best match found so far as the MAP estimate. This approach is quite simple but has a critical drawback. The number of samples needs to be sufficiently large in order to capture all areas of interest (i.e. areas of high probability). Unfortunately, the relative area of high probability with respect to the whole search space shrinks exponentially with the number of dimensions of the search space. And hence, the required number of particles to obtain a reasonable MAP estimate rises exponentially with the dimensionality. This is an effect of the *curse of dimensionality* [9]. The second challenge is sample proposal generation. Intuitively, we want to place the samples in areas of high probability. At the same time, due to limited resources, we want to spare areas of low probability in favor of efficient sample usage. This leads to a chicken-egg-problem, since we do not know the probability distribution a priori (otherwise, MAP inference would become trivial). The Markov chain Monte-Carlo (MCMC) framework presented in Section 3.4 can be applied in such situations. MCMC sampling in high-dimensional space may work out well but requires cleverly designed MCMC moves tailored to the particular computer vision task at hand [122].

The curse of dimensionality problem can be alleviated by applying a *hybrid* approach combining stochastic methods with message passing schemes [62, 61]. The particle belief propagation algorithm (introduced in Sect. 3.5), successfully applied to complex computer vision problems [50, 112, 91, 90], is such a prominent approach. In the message passing framework, sampling is performed over each message or node variable independently. Hence, instead of having one global and high-dimensional MCMC chain, the hybrid approach leads to a large number of low-dimensional (decoupled and parallel) MCMC chains.

Generating sample proposals is easier in low dimensional spaces but still requires a large amount of application-dependent tuning. Hence, many heuristically motivated sampling methods have been proposed, such as random walk proposals [112, 92], neighbor-based sampling [14], and data-driven approaches [91]. Some of them depend on tunable parameters and due to their local generation nature they only provide slow information propagation. In the context of message-passing in hybrid Bayes networks, Koller et al. argue that a sensible choice for good sampling proposals are the marginal distributions [62]. Obviously, this choice is impractical, as the marginal distributions are a priori unknown. As a solution to this dilemma, they propose to use an approximate marginal distribution based on the current message estimate. This results in an iterative scheme in which the message estimates and the approximate marginal distributions are alternately improved. This finding was rediscovered by Ihler and McAllester [50], leading to the sum-product particle belief

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

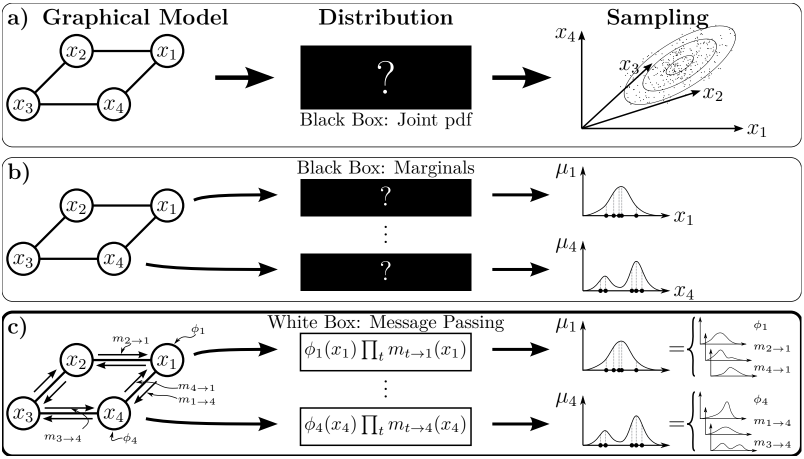


Figure 4.1: Overview of MCMC sampling approaches for inference in probabilistic graphical models: a) Sampling from the joint state space requires an infeasibly large number of particles due to the curse of dimensionality. b) Sampling from the marginal distributions μ (provided by, e.g., belief propagation) is better suited for large-scale problems. However, sampling from μ is treated as a *black box* and requires application-dependent parameter tuning. c) Proposed approach: Same as b) but exploits the structure of μ by combining slice sampling with message-passing. This leads to robust and fast particle sampling without parameter tuning.

propagation (PBP) approach. In [50] it is further proposed to use a Metropolis-Hastings MCMC sampler with Gaussian proposals. A max-product variant for MAP inference was later on proposed in [67].

All the previously mentioned sampling approaches treat the inference task as a black box. They ignore the underlying structure of the marginal distributions. In the following section, a novel sampling method is proposed which is tailored to the max-product message-passing scheme of the particle belief propagation algorithm presented in Sect. 3.5. The proposed approach efficiently generates high-informative sample proposals which respect long-range variable dependencies in the graphical model while requiring significantly less MCMC iterations than black-box Metropolis samplers. Figure 4.1 visualizes the proposed approach (bottom) in comparison to previous methods (top).

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

4.1 Slice-Sampling Particle Max-Product

The intuition behind the proposed approach is to follow the divide-and-conquer principle: instead of trying to sample from the marginal distribution directly, which can be arbitrary complex, the full distribution is decomposed in its single contributing elements and sampling is applied on these much easier shaped, low dimensional distributions.

The proposed method is described step-by-step, starting in Sect. 4.1.1 with the product-max sampler which forms the core sampling method of our approach. In Sect. 4.1.2 our main contribution, sampling from max-marginal distributions in the max-product belief propagation framework, is derived. An empirical random walk analysis with comparison to the baseline PBP algorithm is presented in Sect. 4.1.4. In Sect. 4.2, an extension of our approach is presented. Sampling behavior is further improved by applying two modifications. First, the loopy belief propagation scheme is replaced with a state-of-the-art convergent message-passing algorithm. This increases global consensus of the entire graphical model and hence leads to better guiding of particle sampling. Second, to enforce monotonicity of MAP energy, a particle selection mechanism based on the diverse particle max-product (DPMP) method of [90] is applied. Here, the likelihood to discard particles corresponding to previously discovered high-probable modes is significantly reduced while at the same time preserving diversity in the particle set.

4.1.1 Sampling from Product-Max Distributions

A simple but very important observation is that the (pseudo) max-marginal beliefs which can be computed by max-product message-passing algorithms such as MP-BP (cf. Sect. 3.3.1), TRBP (cf. Sect. 3.3.3), and Metropolis-Hastings particle belief propagation (MH-PBP) (cf. Sect. 3.5) have the following structure:

$$q(x) \propto q_0(x) \prod_{l=1}^{L} \max_{y \in \mathcal{Y}_l} q_l(x, y). \tag{4.1}$$

This distribution is in the remainder of this thesis referred to as a *product-max distribution*. In the following is shown how slice sampling (cf. Sect. 3.4.3) can be applied to draw samples from product-max distributions. Slice sampling has several advantages towards other sampling methods such as Metropolis-Hastings sampling. The method is free of sensitive tuning-parameters and the MCMC chain mixing behavior is much better [88]. The only requirement is that the computation of the slice regions $\{x \; ; \; q_0(x) \geq u\}$ and $\{x \; ; \; q_l(x, y) \geq u\}$ for $l \geq 1$ must be cheap. This is usually not an issue. Many potentials q_0 and q_l are given in closed-form and very often a closed-form (partially) inverse exists. On the other hand, slice regions of more general potentials such as data-driven functions (for example log-linear models

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

with arbitrary feature functions, cf. Eq. (3.14)) can be approximated with moderate costs using, e.g., the step-out procedure (cf. Sect. 3.4.3).

It can be observed that Eq. (4.1) factorizes to $q(x) \propto \prod_{l=0}^{L} q_l(x)$ with

$$q_l(x) = \max_{y \in \mathcal{Y}_l} q_l(x,y) \quad \text{for all } l \geq 1. \tag{4.2}$$

This factorized form is directly suitable for efficient sampling with the product slice sampler (cf. Eq. (3.65)). The product slice sampler requires an efficient way to compute the following slice regions:

$$A_{q_l}(u_l^{\langle m\rangle}) = \{x \; ; \; q_l(x) \geq u_l^{\langle m\rangle}\} = \{x \; ; \; \max_{y \in \mathcal{Y}_l} q_l(x,y) \geq u_l^{\langle m\rangle}\}. \tag{4.3}$$

Assuming discrete and finite state spaces $\mathcal{Y}_l$, one can in Eq. (4.3) factor out the max-operation over $\mathcal{Y}_l$ which turns into a union over $\mathcal{Y}_l$:

$$A_{q_l}(u_l^{\langle m\rangle}) = \bigcup_{y \in \mathcal{Y}_l} \{x \; ; \; q_l(x,y) \geq u_l^{\langle m\rangle}\} = \bigcup_{y \in \mathcal{Y}_l} A_{q_l}(u_l^{\langle m\rangle}, y) \tag{4.4}$$

with $A_{q_l}(u,y) = \{x \; ; \; q_l(x,y) \geq u\}$. The equivalence of (4.3) and (4.4) can be easily seen by the following set-theoretic equivalences:

$$\begin{aligned}
z \in \{x \; ; \; \max_{y \in \mathcal{Y}_l} q_l(x,y) \geq u_l^{\langle m\rangle}\} &\Leftrightarrow \max_{y \in \mathcal{Y}_l} q_l(z,y) \geq u_l^{\langle m\rangle} && \text{(4.5a)}\\
&\Leftrightarrow \exists y \in \mathcal{Y}_l \;:\; q_l(z,y) \geq u_l^{\langle m\rangle} && \text{(4.5b)}\\
&\Leftrightarrow \exists y \in \mathcal{Y}_l \;:\; z \in \{x \; ; \; q_l(x,y) \geq u_l^{\langle m\rangle}\} && \text{(4.5c)}\\
&\Leftrightarrow z \in \bigcup_{y \in \mathcal{Y}_l} \{x \; ; \; q_l(x,y) \geq u_l^{\langle m\rangle}\}. && \text{(4.5d)}
\end{aligned}$$

According to Eq. (3.67), the slice regions $A_{q_l}(u_l^{\langle m\rangle})$ are combined to a *joint* slice region via intersection: $A_q(u^{\langle m\rangle}) = \bigcap_l A_{q_l}(u_l^{\langle m\rangle})$. Bringing everything together leads to the following relationship between product-max distributions and its corresponding slice regions:

$$\begin{array}{cccccc}
q(x) & \propto & q_0(x) & \prod_{l=1}^{L} & \max_{y \in \mathcal{Y}_l} & q_l(x,y) \\
\updownarrow & & \updownarrow & & & \updownarrow \\
A_q(u^{\langle m\rangle}) & = & A_{q_0}(u_0^{\langle m\rangle}) & \bigcap_{l=1}^{L} & \bigcup_{y \in \mathcal{Y}_l} & A_{q_l}(u_l^{\langle m\rangle}, y)
\end{array} \tag{4.6}$$

The correspondence between the operators is highlighted with colors where the $\prod \leftrightarrow \bigcap$ correspondence follows from the product slice sampler (cf. Eqs. (3.65)–(3.67)) and the $\max \leftrightarrow \bigcup$ correspondence is shown above in Eqs. (4.5). The product-max sampler is summarized in Alg. 5. In case that the slice regions for $q_0(x)$ or $q_l(x,y)$ do not have closed-form solutions (as is the case when using, e.g, data-driven likelihood functions), one can resort to approximating the slice regions. This then requires an additional acceptance test

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

Algorithm 5 Product-max sampler

```
Input: Initial particle $x^{\langle 0\rangle}$, slice regions $A_{q_l}(u, y)$ for each factor $q_l$, discrete and finite
    sets $\mathcal{Y}_l$
Ensure: $x \sim \prod_{l=1}^{L} \max_{y\in\mathcal{Y}_l} q_l(x, y)$
 1: for $m = 1$ to $M$ do
 2:     for $l = 1$ to $L$ do
 3:         Sample $u_l^{\langle m\rangle} \sim \mathcal{U}\,[0, q_l(x^{\langle m-1\rangle})]$
 4:         Compute region $A_{q_l}(u_l^{\langle m\rangle}) = \bigcup_{y\in\mathcal{Y}_l} A_{q_l}(u_l^{\langle m\rangle}, y)$
 5:     end for
 6:     Compute region $A_q(u^{\langle m\rangle}) = \bigcap_{l=1}^{L} A_{q_l}(u_l^{\langle m\rangle})$
 7:     Sample new particle $x^{\langle m\rangle} \sim \mathcal{U}(A_q(u^{\langle m\rangle}))$
 8:     \\ (Optional) Acceptance test for approximated slice regions:
 9:     if $\forall l \ : \ q_l(x^{\langle m\rangle}) \geq u_l^{\langle m\rangle}$ then
10:         Accept $x^{\langle m\rangle}$
11:     else
12:         \\ Shrink slice (cf.  Sect. 3.4.3):
13:         if $x^{\langle m\rangle} < x^{\langle m-1\rangle}$ then
14:             $A_q(u^{\langle m\rangle}) \leftarrow A_q(u^{\langle m\rangle}) \cap [x^{\langle m\rangle}, \infty)$
15:         else
16:             $A_q(u^{\langle m\rangle}) \leftarrow A_q(u^{\langle m\rangle}) \cap (-\infty, x^{\langle m\rangle}]$
17:         end if
18:         goto 7
19:     end if
20: end for
21: Set $x = x^{\langle M\rangle}$
```

4.1.2 Particle Max-Product

In this section, a novel particle sampling strategy is proposed which exploits the message-passing structure of the underlying inference framework for generating high-informative particles. The core inference method is the PBP approach of Sect. 3.5. In the literature, particles for PBP are generated either from Metropolis MCMC simulations or from heuristic proposals. In both cases, parameter tuning may have a high impact on the quality of the generated particles. The proposed approach in this chapter exploits the properties of (product) slice sampling which is robust against parameter selection.

The following derivation assumes pairwise graphical models. The generalization to higher-order models is straightforward. Similar to PBP introduced in Sect. 3.5, particles are drawn from max-marginal distributions. But the strategy to generate them is different. The Metropolis-Hastings is replaced by the product-max sampler

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

introduced in the previous section. The pseudo max-marginals computed by the (loopy) belief propagation algorithm have the following structure (cf. Sect. 3.5):

$$\hat{\mu}_s(x_s) = \phi_s(x_s) \prod_{t \in \mathcal{N}_s} \max_{x_t \in \mathcal{P}_t} [\phi_t(x_t)\phi_{s,t}(x_s, x_t) \prod_{r \in \mathcal{N}_t \setminus \{s\}} \hat{m}_{r \to t}(x_t)]. \tag{4.7}$$

This equation follows by substituting the message $\hat{m}_{t \to s}(x_s)$ defined by Eq. (3.69) into the belief equation (3.70). Note that this distribution is in fact a product-max distribution over x_s as visually highlighted corresponding to Eq. (4.6). Hence, Alg. 5 can be directly applied to sample from the pseudo max-marginals $\hat{\mu}_s(x_s)$. The relationship between Eq. (4.1) and Eq. (4.7) can be seen by using the following substitutions:

$$L = |\mathcal{N}_s| \tag{4.8a}$$

$$q_0(x_s) = \phi_s(x_s) \tag{4.8b}$$

$$\mathcal{Y}_l = \mathcal{P}_{t^{(l)}} \tag{4.8c}$$

$$q_l(x_s, x_t) = \phi_t(x_t)\phi_{s,t}(x_s, x_t) \prod_{r \in \mathcal{N}_t \setminus \{s\}} \hat{m}_{r \to t}(x_t). \tag{4.8d}$$

The corresponding slice regions A_{q_l} are thus defined as follows:

$$A_{q_0}(u) = A_{\phi_s}(u) \tag{4.9a}$$

$$A_{q_l}(u) = A_{\phi_{s,t}}(\hat{u}, x_t), \quad \hat{u} = u \cdot \left[\phi_t(x_t) \prod_{r \in \mathcal{N}_t \setminus \{s\}} \hat{m}_{r \to t}(x_t)\right]^{-1} \tag{4.9b}$$

with $A_{\phi_s}(u) = \{x_s \ ; \ \phi_s(x_s) \geq u\}$ and $A_{\phi_{s,t}}(\hat{u}, x_t) = \{x_s \ ; \ \phi_{s,t}(x_s, x_t) \geq \hat{u}\}$.

In the following we switch to the negative logarithm representation of the beliefs, messages, and factor potentials, as it is common practice for increasing numerical stability. This essentially leads to replacing products with sums and max-operators with min-operators. The proposed method is summarized in Alg. 6. Note that the set-operators union and intersection of Alg. 5 are not altered, since the slice regions stay the same in both the max-product and min-sum variants. The slice regions for unary potentials $A_{\phi_s}(u) = \{x_s \ ; \ \phi_s(x_s) \geq u\}$ transforms to $\overline{A}_{\phi_s}(\bar{u}) := \{x_s \ ; \ \psi_s(x_s) \leq \bar{u}\}$ where $\bar{u} := -\log(u)$. For binary and higher-order potentials the definitions apply analogously.

The above described algorithm, summarized in Alg. 6, is in the following referred to as slice-sampling particle belief propagation (S-PBP).

Convergence Properties The core of the proposed approach is the product slice sampler which introduces auxiliary variables $\{u_l\}_{l=1,\dots,L}$ for each factor respectively. Neal [88] mentions in his comparison of single auxiliary variable samplers to multiple auxiliary variable samplers that the MCMC convergence time for product slice sampling is in $\mathcal{O}(L)$. This implicates that samples are more correlated with larger L

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

Algorithm 6 Slice-Sampling Particle Belief Propagation

```
Input: Initial set of particles: $\{x_s^{(i)}\}_{i=1,\dots,p}$
Initialize the messages $M^0_{t\to s}(x_s) = 0$ and log disbelief $B^0_s(x_s^{(i)}) = 0\ \forall s,t$
for BP iteration $n = 1$ to $N$ do
    for each node $s$ and each particle $i = 1,\dots,p$ do
        Initialize sampling chain $x_s^{(i)\langle 0\rangle} \leftarrow x_s^{(i)}$
        for MCMC iteration $m = 1,\dots,M$ do
            Sample $u_0 \sim \mathcal{U}[0,1]$
            $\bar{u}_0 \leftarrow \psi_s(x_s^{(i)\langle m-1\rangle}) - \log(u_0)$
            for $t \in \mathcal{N}_s$ do
                Sample $u_t \sim \mathcal{U}[0,1]$
                $\bar{u}_t \leftarrow M^n_{t\to s}(x_s^{(i)\langle m-1\rangle}) - \log(u_t)$
                Compute region $\overline{A}_{M^n_{t\to s}}(\bar{u}_t) = \bigcup_{x_t \in \mathcal{P}_t} \overline{A}_{\psi_{s,t}}(\bar{u}_t - \hat{M}_{t\to s}(x_t), x_t)$
            end for
            Compute region $\overline{A}_{B_s}(\bar{u}) = \overline{A}_{\psi_s}(\bar{u}_0) \bigcap_{t \in \mathcal{N}_s} \overline{A}_{M_{t\to s}}(\bar{u}_t)$
            Sample new particle candidate $\bar{x}_s^{(i)\langle m\rangle} \sim \mathcal{U}(\overline{A}_{B_s}(\bar{u}))$
            Calc. belief $B^n_s(\bar{x}_s^{(i)\langle m\rangle})$ from Eqs. (3.70), (3.69)
            if $\psi_s(\bar{x}_s^{(i)\langle m\rangle}) \le \bar{u}_0$ and $M^n_{t\to s}(\bar{x}_s^{(i)\langle m\rangle}) \le \bar{u}_t\ \forall t \in \mathcal{N}_s$ then
                Accept: $x_s^{(i)\langle m\rangle} \leftarrow \bar{x}_s^{(i)\langle m\rangle}$
            end if
        end for
        $x_s^{(i)} \leftarrow x_s^{(i)\langle M\rangle}$
    end for
    Normalize messages and beliefs
end for
```

and therefore has a negative impact on MCMC chain mixing. In our case, the number of factors is limited by the structure of the underlying graphical model $\mathcal{G}$. To be more precise, $L_s = 1 + |\mathcal{N}_s|$, where $|\mathcal{N}_s|$ is the number of neighbors of node s. Since MCMC is performed for each node independently and can be done in parallel, the overall convergence time for the complete graph is of order $L_{\max} = 1 + \max_{s\in\mathcal{V}} |\mathcal{N}_s|$. Note that $L_{\max}$ is typically < 10 (e.g. $L_{\max} = 3$ for chain graphs and $L_{\max} = 5$ for grid structures with a 4-neighborhood).

Computational Complexity The computational complexity can be a useful cue in comparing different algorithms. A proxy for the computational complexity are the sum of number of *factor potential evaluations* and the number of *slice computations*. This is a meaningful measure, because the two operations provide the smallest possible interface between the graphical model representation and the inference algorithms.

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

The baseline method is MH-PBP as introduced in Sect. 3.5. The computational complexity for MH-PBP is $\mathcal{O}(NSpM\ (1+Vp))$ and for S-PBP is $\mathcal{O}(NSpM(3+2Vp))$ given the number of PBP iterations N, nodes S, particles p, MCMC iterations M and the average number of neighbors per node V. This indicates a doubling of computation time of S-PBP compared to MH-PBP which is due to the overhead introduced for computing the interval regions $A(u)$. Note, however, that the MCMC sampling chain can be significantly shorter in S-PBP than in MH-PBP due to much better chain mixing properties of slice sampling. The chain mixing behavior is analyzed in more detail in Sect. 4.1.4.

Multivariate Distributions To deal with multivariate label spaces, i.e. $\mathcal{L}_s \in \mathbb{R}^d$ for $d > 1$, there are several possible choices. One approach is to randomly select one dimension in each MCMC step and slice sample on this dimension while the other dimensions are held fixed. Another choice is to apply Gibbs sampling. This is a deterministic variation of the former method where over each dimension is sampled exactly once in an (arbitrarily chosen) predefined order.

4.1.3 Computing the Slice Regions

The proposed slice sampling approach relies on fast computation of the slice regions. This section summarizes approaches for computing either exact or approximate slice regions.

Exact slice regions can be provided when the potential function under test f is (partially) invertible with a closed-form representation. With slight abuse of convention we refer to these functions in short as *invertible* functions and to all other types as *non-invertible* or *data-driven* functions. Computing the exact slice region $\overline{A}_f(\bar{u}) = \{x;\ f(x) \leq \bar{u}\}$ then involves two steps. First, determining a partitioning of the domain of f into *branches* $\{D_i\}_i$ and pivot points $\{p_i\}_i$ in which $f(x)$ is monotonic decreasing for all $x \in D_i$ with $x < p_i$ and monotonic increasing for all $x \in D_i$ with $x \geq p_i$. Second, computing the inverses $f_{i,L}^{-1}$ and $f_{i,R}^{-1}$ of f for each branch i corresponding to the left and right side of p_i correspondingly. Finally, the slice region $\overline{A}_f(\bar{u})$ is a union over all intervals $[f_{i,L}^{-1}(\bar{u}), f_{i,R}^{-1}(\bar{u})]$ for which $\bar{u}$ is in the image of f restricted to branch D_i (or equivalently in the domain of $f_{i,L}^{-1}$ and $f_{i,R}^{-1}$).

For example, the function $f(x) = x^2$ has only one branch $D_1 = (-\infty, \infty)$ with pivot point $p_1 = 0$ and partial inverses $f_{1,L}^{-1}(y) = -\sqrt{y}$ and $f_{1,R}^{-1}(y) = \sqrt{y}$. The corresponding slice region is $\overline{A}_f(\bar{u}) = \emptyset$ if $\bar{u} < 0$ and $\overline{A}_f(\bar{u}) = [-\sqrt{y}, \sqrt{y}]$ otherwise.

The process for defining $A_{\psi_s}(u)$ and/or $A_{\psi_{s,t}}^{x_t}(u)$ as described above can be automated using standard computer algebra solvers. A MATLAB®-MuPAD® interface was developed to solve the inequalities and derive the slice regions automatically. This avoids tedious manual derivations which can get quite large and are likely to contain mistakes.

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

Approximate Slice Regions

Closed-form solutions for the slice regions are not always available. The proposed framework is capable of handling such a case. This works by first computing an approximation $A_f^{\text{approx}}(u)$ of the true slice region and perform rejection sampling with slice shrinking. Neal [88] proposes two methods for univariate functions: the step-out procedure and the doubling procedure (cf. Sect. 3.4.3). The two methods successively enlarge an initial small slice region until a stopping criterion is reached. This requires repeated factor potential evaluation, which has, depending on the selected step size and the shape of the factor potentials, a negative impact on the algorithm runtime.

Another approach is to directly derive an outer bound approximation $A_f^{\text{approx}}(u) \supset A_f(u)$. Provided that the state space is bounded with $\mathcal{X} \subset [l{,}r]$ and $-\infty < l < r < \infty$, a naive outer bound approximation is $A_f^{\text{approx}}(u) = [l{,}r]$. A downside of such a loose outer bound approximation is that the number of sample rejections in the slice shrinking step rapidly increases, which, again, increases algorithm runtime. In practice, the slow-down is not as large. Recall that samples are drawn from the max-marginal distribution $\mu_s(x_s)$. This involves the computation of slice regions for the unary potential $A_{\phi_s}(u)$ and *all* pairwise potentials $A_{\phi_{s,t}}(\hat{u}, x_t)$ to neighboring nodes $t \in \mathcal{N}_s$. Further recall from Alg. 7 that the region from which a new sample is drawn is the *intersection* of multiple slice regions. Assume that at least one slice region can be computed exactly and the resulting slice region area is reasonably small compared to the sample space. Due to the intersection operator, the area of the resulting sampling region is then at most as large as the smallest slice region.

If, for example, the slice region for the unary potential $\phi_s(x_s)$ is approximated as $(-\infty, \infty)$ and the exact (bounded) slice regions for the pairwise potentials are provided, then the intersection of all regions is bounded. Omitting the slice shrinkage step would in this case lead to sampling from the distribution $\prod_{t \in \mathcal{N}_s} \hat{m}_{t \to s}(x_s)$ instead of the approximate max-marginal distribution $\phi_s(x_s) \prod_{t \in \mathcal{N}_s} \hat{m}_{t \to s}(x_s)$, hence ignoring the influence of the unary potential $\phi_s(x_s)$ at all. Therefore, slice shrinking is only triggered when the unary potential $\phi_s(x_s)$ *disagrees* with the rest of the network.

Further speedup in product slice sampling can be achieved by summarizing all non-invertible terms into a single factor and performing slice approximation solely for this summarized factor. This reduces the number of multiple slice approximation steps to a single slice approximation and furthermore reduces the number of set union and intersection operations over the slice regions. In the extreme case when all potential functions are non-invertible this method reduces to black-box sampling (cf. Fig. 4.1).

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

Slice Approximation for Gaussian Mixture Models

In the following it is shown that an outer bound approximation can be provided when the factor potential is a Gaussian mixture model:

$$\phi_s(x_s) = \sum_{k=1}^{K} w_k \mathcal{N}(x_s \mid \mu_k, \sigma_k^2) \tag{4.10}$$

with component means μ_k, component variances σ_k^2 and component weights $w_k \geq 0$, where $\sum_k w_k = 1$. An outer bound approximation of $A_{\phi_s}(u)$ can be constructed by relaxing $\phi_s(x_s)$ as follows

$$\phi_s^{\text{approx}}(x_s) = \max_{k=1,\ldots,K} K w_k \mathcal{N}(x_s \mid \mu_k, \sigma_k^2) \geq \phi_s(x_s). \tag{4.11}$$

This can be easily seen by the following inequality

$$\begin{aligned} \sum_{k=1}^{K} w_k \mathcal{N}(x \mid \mu_k, \sigma_k^2) &\leq \sum_{k=1}^{K} \max_{k'=1,\ldots,K} w_{k'} \mathcal{N}(x \mid \mu_{k'}, \sigma_{k'}^2) \\ &= K \max_{k'=1,\ldots,K} w_{k'} \mathcal{N}(x \mid \mu_{k'}, \sigma_{k'}^2) = \phi_s^{\text{approx}}(x_s). \end{aligned} \tag{4.12}$$

The corresponding slice region has the following form

$$A_{\phi_s}^{\text{approx}}(u) = \cup_{k=1}^{K} [\mu_k - d_{2,k}, \mu_k + d_{2,k}], \tag{4.13}$$

with $d_k = \sigma_k \sqrt{2(\bar{u} + \log(K w_k) - \log(\sqrt{2\pi}\sigma_k))}$.

Note that other approximations are also possible, e.g., $\max_k \mathcal{N}(x_s \mid \mu_k, \sigma_k^2)$, but were not considered in the remainder of this thesis.

Table 4.1 provides a summary of the discussed potential functions and their corresponding slice regions or slice region approximations.

4.1.4 Random Walk Analysis

In the following, the chain mixing behavior of the proposed S-PBP algorithm is analyzed on the application of edge-preserving image denoising. Each pixel is associated to a random variable x_s, where $s = 1, \ldots, W \cdot H$ with an image of W columns and H rows. We restrict ourselves to grayscale images. The observed variables $d_s \in [0,1]$ take the (noisy) intensity values of the corresponding image pixels. The random variables x_s represent the intensities of the true (denoised) image pixels. Assuming i.i.d. Gaussian noise and a robust smoothness prior between neighbored pixels, the image denoising model is as follows:

$$\begin{aligned} \psi_s(x_s) &= \theta_1 (x_s - d_s)^2, \\ \psi_{s,t}(x_s, x_t) &= \theta_2 \min\{\theta_3, (x_s - x_t)^2\}. \end{aligned} \tag{4.14}$$

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

Table 4.1: Listing of frequently used potential functions and its corresponding (approximate) slice regions.

Type	Potential $\phi_s(x)$	Slice region $A_{\phi_s}(u)$	
Gaussian	$\exp\big[-\frac{(x_s-\mu)^2}{2\sigma^2}\big]$	$[\mu-\sigma\sqrt{2\bar{u}},\mu+\sigma\sqrt{2\bar{u}}]$	
Truncated quadratic	$\exp\big[-\min\{\delta,\frac{(x_s-\mu)^2}{2\sigma^2}\}\big]$	$\bar{u}<\delta \;:\; [\mu-\sigma\sqrt{2\bar{u}},\mu+\sigma\sqrt{2\bar{u}}]$ $\bar{u}\geq\delta \;:\; \mathbb{R}$	
Absolute deviation	$\exp\big[-\lambda\lvert x_s-\mu\rvert\big]$	$[\mu-\frac{\bar{u}}{\lambda},\mu+\frac{\bar{u}}{\lambda}]$	
Robust penalty	$\exp\big[-\rho((x_s-\mu)^2)\big]$	$[\mu-\sqrt{\rho^{-1}(\bar{u})},\mu+\sqrt{\rho^{-1}(\bar{u})}]$	
Data driven	$f(x_s)$	step-out method	slice shrinking
Data driven (bounded)	$f(x_s) \;:\; x_s\in\mathcal{X}_s\subset[l,r]$	$\mathcal{X}_s$	
Gaussian mixture	$\sum\limits_k\mathcal{N}(x_s\mid\theta_k)$	$\bigcup\limits_k[\mu-d_k,\mu+d_k]$	

This model is suitable for our analysis due to its moderate problem size, allowing an empirical evaluation with hundreds of runs. Yet, it has sufficient complexity in terms of a high number of highly-correlated random variables and a graph structure with many tight loops.

For minimizing particle noise in the final estimation result an annealing scheme is used where the target belief distribution is modified to $b_s^n(x_s^{(i)})^{1/T_n}$, where

$$T_n = T_0\cdot(T_N/T_0)^{n/N} \tag{4.15}$$

is the temperature at PBP iteration n, T_0 is the start temperature, and T_N the end temperature. Given this annealing scheme the temperature is successively reduced for each new iteration n.

The evaluation was done on an example image as shown in Fig. 4.2. The training and testing sets each include 10 noisy image instances with Gaussian noise standard deviation $\sigma = 0.05$. Training of the parameter vector $\boldsymbol{\theta} = \{\theta_1,\theta_2,\theta_3\}$ is done by minimizing the *empirical risk* $R(\boldsymbol{\theta}) = \frac{1}{K}\sum_{i=1}^{K} L(\mathbf{x}_{\boldsymbol{\theta}}^{(i)},\mathbf{y}^{(i)})$ given the *loss function* $L(\mathbf{x},\mathbf{y}) = \|\mathbf{x}-\mathbf{y}\|_2^2$ where $\{\mathbf{y}^{(i)},\mathbf{d}^{(i)}\}$ is the training data pair with ground-truth $\mathbf{y}^{(i)}$ and noisy observation $\mathbf{d}^{(i)}$, and $\mathbf{x}_{\boldsymbol{\theta}}^{(i)}$ is the MAP estimate given $\mathbf{d}^{(i)}$ and the parameter $\boldsymbol{\theta}$. Learned parameters are $\theta_1 = 0.756$, $\theta_2 = 1.170$, $\theta_3 = 0.0059$.

Comparing S-PBP with MH-PBP The efficiency of slice sampling is compared to the baseline Metropolis-Hastings sampling approach. The experimental setup

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte
Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

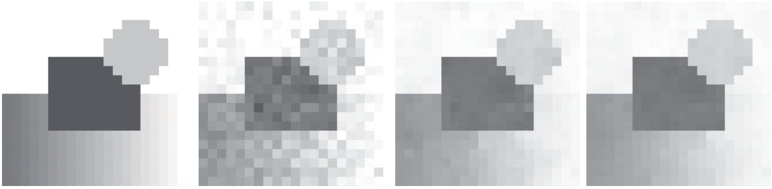


Figure 4.2: Denoising example: Groundtruth (left), noisy input example (middle left), reconstruction with MH-PBP (middle right), reconstruction with our proposed S-PBP method (right).

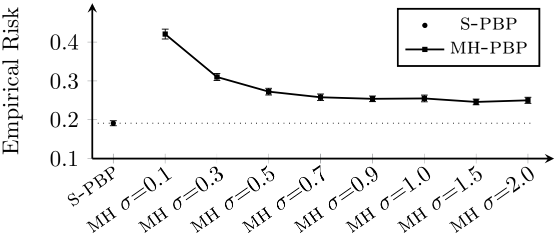


Figure 4.3: Comparison of the empirical risk for S-PBP and MH-PBP with different proposal distributions.

has the following parameters: $N = 100$ PBP iterations, $p = 5$ particles, and a temperature schedule of $T_0 = 1$ to $T_N = 10^{-4}$. An MCMC chain of $M = 500$ samples was generated for each particle and in each PBP iteration. The iteration numbers are chosen to be more than sufficiently large in order to guarantee convergence and to collect statistical information in the MCMC chains in steady-state situations. For the MH-PBP proposal distribution the family of Gaussian distributions $p_\sigma(x \mid x^{\langle m-1 \rangle}) = (2\pi\sigma^2)^{-0.5} \cdot \exp[-0.5(x - x^{\langle m-1 \rangle})^2 \cdot \sigma^{-2}]$ is used. In order to provide a fair comparison the proposal distribution is adapted to the current temperature by using $p_\sigma(x \mid x^{\langle m-1 \rangle})^{1/T_n}$ instead.

Figure 4.3 shows a comparison of the empirical risk for different MH-PBP proposal distributions. For $\sigma > 0.7$ the empirical risk stays nearly at the same level and thus $\sigma = 0.7$ was selected for all further experiments. Another observation is that S-PBP outperforms MH-PBP in terms of minimal empirical risk. This is because the reconstructed images with MH-PBP have always much higher noise than images reconstructed with S-PBP. This effect can be significantly reduced by averaging over particles instead of only selecting the best one as stated in Eq. (3.71).

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

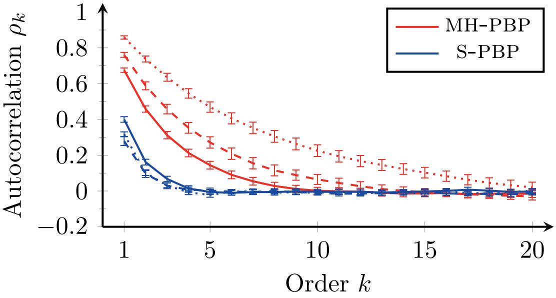


Figure 4.4: Comparison of S-PBP and MH-PBP at different PBP iterations (dotted $n = 30$, dashed $n = 50$, and solid $n = 70$) using an annealing schedule.

For comparing the random walk behavior of the MCMC sampling chains from S-PBP and MH-PBP, the normalized autocorrelation function

$$\rho_k = \frac{\sum_{m=1}^{M-k} (x^{\langle m \rangle} - \bar{x})(x^{\langle m-k \rangle} - \bar{x})}{\sum_{m=1}^{M-k} (x^{\langle m \rangle} - \bar{x})^2}, \tag{4.16}$$

where $\bar{x} = \frac{1}{M} \sum x^{\langle m \rangle}$, is used [118]. Only the last 50 % of the MCMC chain is considered to skip any burn-in phase. Figure 4.4 shows a comparison of the first 20 k-th order autocorrelation of S-PBP and MH-PBP at different PBP iterations n (and thus at different temperatures T_n). It can be observed that the MH-PBP method produces a much higher autocorrelation than the S-PBP method, thus it can be concluded that the MCMC chain mixing behavior of S-PBP outperforms MH-PBP.

The evaluation in this section only focused on synthetic data. An evaluation with real-world data is provided in Sect. 5.1 with the application of part-based object tracking.

4.2 Tree-Reweighted Particle Max-Product

In the previous section, the inference mechanism of S-PBP is based on loopy belief propagation. As already pointed out in Sect. 3.3.2, loopy belief propagation has poor convergence behavior. Convergence is not guaranteed. Non-convergence often manifests in oscillating message updates from which the algorithm must be aborted after an arbitrary number of maximally allowed iterations. In such cases, the computed pseudo max-marginals can be misleading which has a negative impact on the particle sampling quality.

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

It is shown in the following that the proposed product-max sampling is also applicable on other message-passing methods which have much better convergence behavior. The derivation focuses on the TRBP algorithm with activated message-damping as introduced in Sect. 3.3.3. In preliminary tests it was observed that this approach achieves sufficiently fast convergence.

The pseudo max-marginals of TRBP have the following structure:

$$\hat{\mu}_s(x_s) \propto \phi_s(x_s) \prod_{t\in\mathcal{N}_s} \underbrace{\left(\max_{x_t\in\mathcal{P}_t} \phi_{s,t}(x_s,x_t)^{\alpha}\,\overline{m}_{t\to s}(x_t)^{\alpha\rho_{st}}\right) m^{\text{old}}_{t\to s}(x_s)^{\rho_{st}(1-\alpha)}}_{m_{t\to s}(x_s)^{\rho_{ts}}} \tag{4.17}$$

This equation follows by substituting the damped message $m_{t\to s}(x_s)$ as defined in Eq. (3.27) into the belief equation (3.26) where $\overline{m}_{t\to s}(x_t)$ is the premessage and $m^{\text{old}}_{t\to s}(x_s)$ is the old message from the previous iteration. One can observe that this form almost perfectly matches the pattern required by the product-max sampler in Eq. (4.6) of Sect. 4.1.1. The only term which does not match the product-max pattern is $m^{\text{old}}_{t\to s}(x_s)^{\rho_{st}(1-\alpha)}$ originating from the message damping mechanism. This term is recursively defined through the iterative message passing mechanism. A repeated substitution by its recursive definition is not desirable because the number of factors L for the product slice sampler would grow with each substitution. As pointed out in Sect. 4.1.2, larger L requires more MCMC iterations and thus has a negative impact on the computational overhead. Therefore, this approach is not tractable. Another approach is to slightly modify the target distribution (4.17) by replacing $m_{t\to s}(x_s)^{\rho_{ts}}$ with $\left({m_{t\to s}(x_s)}/{m^{\text{old}}_{t\to s}(x_s)^{1-\alpha}}\right)^{\frac{\rho_{ts}}{\alpha}}$, leading to the following tractable proposal distribution:

$$q_s(x_s) \propto \phi_s(x_s) \prod_{t\in\mathcal{N}_s} \left(\frac{m_{t\to s}(x_s)}{m^{\text{old}}_{t\to s}(x_s)^{1-\alpha}}\right)^{\frac{\rho_{ts}}{\alpha}} \tag{4.18}$$

$$= \phi_s(x_s) \prod_{t\in\mathcal{N}_s} \max_{x_t\in\mathcal{P}_t} \phi_{s,t}(x_s,x_t)\,\overline{m}_{t\to s}(x_t)^{\rho_{st}} \tag{4.19}$$

In the case of disabled message damping ($\alpha = 1$), the pseudo max-marginal in Eq. (4.17) reduces to

$$\hat{\mu}_s(x_s) \propto \phi_s(x_s) \prod_{t\in\mathcal{N}_s} \max_{x_t\in\mathcal{P}_t} \phi_{s,t}(x_s,x_t)\,\overline{m}_{t\to s}(x_t)^{\rho_{st}} \propto q_s(x_s). \tag{4.20}$$

That is, the pseudo max-marginal is equivalent up to a scaling factor to the proposal $q_s(x_s)$. In the case of enabled message damping ($\alpha \neq 1$), one can observe the following. If the TRBP algorithm converged to a fixpoint, i.e. $m_{t\to s}(x_s) = m^{\text{old}}_{t\to s}(x_s)$, the proposal distribution equals the pseudo max-marginal $q_s(x_s) = \hat{\mu}_s(x_s)$. The proposed proposal $q_s(x_s)$ in Eq. (4.18) does *not* equal the pseudo max-marginal if both $\alpha \neq 1$ and the TRBP algorithm is not converged. That means that the use

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

Algorithm 7 TRBP max-marginal sampler.

```
Input: Initial particle $x_s^{\langle 0\rangle}$ for vertex $s$, slice regions $A_{\phi_s}(u)$ and $A_{\phi_{s,t}}(u, x_t)$
Ensure: $x_s \sim \hat{\mu}_s(x_s)$
for $m = 1$ to $M$ do
    Sample $u_0^{\langle m\rangle} \sim \mathcal{U}\,[0, \phi_s(x_s^{\langle m-1\rangle})]$
    for $t \in \mathcal{N}_s$ do
        Sample $u_t^{\langle m\rangle} \sim \mathcal{U}\,[0, ({}^{m_{t\to s}(x_s^{\langle m-1\rangle})}/_{m^{\text{old}}_{t\to s}(x_s^{\langle m-1\rangle})^{1-\alpha}})^{\frac{\rho_{ts}}{\alpha}}]$
        Compute region $A_{m_{t\to s}}(u_t^{\langle m\rangle}) = \bigcup_{x_t \in \mathcal{P}_t} A_{\phi_{s,t}}({}^{u_t^{\langle m\rangle}}/_{\overline{m}_{t\to s}(x_t)^{\rho_{st}}}, x_t)$
    end for
    Compute region $A_{q_s}(u^{\langle m\rangle}) = A_{\phi_s}(u_0^{\langle m\rangle}) \bigcap_{t \in \mathcal{N}_s} A_{m_{t\to s}}(u_t^{\langle m\rangle})$
    Sample new particle $x^{\langle m\rangle} \sim \mathcal{U}(A_{q_s}(u^{\langle m\rangle}))$
end for
Set $x_s = x_s^{\langle M\rangle}$
```

of the proposed proposal could potentially have a negative impact on the particle quality in the starting phase of TRBP with enabled message damping. On the other hand, when TRBP converges, the proposal $q_s(x_s)$ successively approaches $\hat{\mu}_s(x_s)$ and thus the particle quality increases from iteration to iteration.

The TRBP max-marginal sampler is summarized in Alg. 7. Note that this method can be interpreted as a generalization of the S-PBP approach of the previous section. Loopy belief propagation is achieved by setting the TRBP parameters $\rho_{st} = 1$ and the damping factor $\alpha = 1$ (no damping).

4.2.1 Diverse Particle Selection

The S-PBP approach presented in the previous sections focuses on improving the MCMC sampling mechanism. However, one drawback of the MCMC PBP framework is that the set of particles is replaced after each sampling step. Already discovered good MAP estimates are often discarded in the next PBP sampling iteration. There are several ways of dealing with this problem: The easiest way is to keep track of the configuration which achieves the lowest energy (highest probability) so far. This guarantees a monotonically decreasing upper bound of the MAP energy. A second approach is to perform temperature annealing. This method is for example applied in the image denoising in the previous section (cf. Sect. 4.1.4) and in the part-based object tracker in the next chapter (cf. Sect. 5.1). Although this approach works reasonably well, it has some downsides. First, another tuning parameter, the annealing speed, is introduced. Second, the risk of getting stuck in local optima is increased, especially when the cooling-down schedule is too rapid.

Other approaches try to keep the most promising particles by first augmenting the old particle set by newly sampled particles, leading to a much larger *augmented*

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

particle set. The augmented particle then must be reduced by applying a filtering step. A simple filter is to discard particles with the smallest max-marginals [14]. This approach tends to cluster the particles to a single mode [90]. The approach of Pacheco et al. [91, 90] follow a similar approach but with another objective than the max-marginals in order to put emphasis towards *diversity*. They showed substantial improvements in MAP estimation on complex problems with multiple modes such as human pose estimation [91], optical flow estimation, and protein side chain prediction [90]. This method is denoted as diverse particle max-product (DPMP).

The criterion for selecting diverse particles in DPMP is to minimize the approximation error of the max-product messages. Since the true max-product messages are not available, they use the messages computed over the augmented particle set as reference. Let $\mathcal{P}_t^{\text{init}}$ denote the initial particle set for vertex t containing $L_t = |\mathcal{P}_t^{\text{init}}|$ particles and $\mathcal{P}_t^{\text{aug}} = \mathcal{P}_t^{\text{init}} \cup \mathcal{P}_t^{\text{new}}$ the particle set augmented with K_t newly generated particles $\mathcal{P}_t^{\text{new}}$. Hence, $|\mathcal{P}_t^{\text{aug}}| = L_t + K_t$. Then, the goal is to filter a subset of L_t particles $\mathcal{P}_t$ from the augmented particle set $\mathcal{P}_t^{\text{aug}}$ according to the following objective:

$$\begin{aligned} \min_{\mathcal{P}_t \subset \mathcal{P}_t^{\text{aug}}} \quad & \sum_{s \in \mathcal{N}_t} \| m_{t\to s}(\mathcal{P}_t) - m_{t \to s}(\mathcal{P}_t^{\text{aug}}) \|_1 \\ \text{s.t.} \quad & |\mathcal{P}_t| = L_t, \end{aligned} \tag{4.21}$$

where $m_{t\to s}(A)$ are message(-vectors) calculated over its corresponding particle set A:

$$m_{t\to s}(A) = \left\{ \max_{x_t \in A} \phi_{s,t}(x_s, x_t) \overline{m}_{t\to s}(x_t) \right\}_{x_s \in \mathcal{P}_s^{\text{aug}}} \tag{4.22}$$

and $\| \cdot \|_1$ is the L1-norm with $\|\mathbf{a}\|_1 = \sum_i |a_i|$. The problem (4.21) is approximately solved with an efficient lazy greedy approach [90].

Figure 4.5 shows a comparison of the temperature annealing approach and diverse particle selection applied on the image denoising example in Sect. 4.1.4. Temperature annealing does not guarantee monotonic decreasing MAP estimates. DPMP with slice-sampling converges much faster.

The particles $\mathcal{P}_t^{\text{new}}$ are generated from heuristic proposal generators that include random-walk, neighbor-based, and data driven proposal distributions [90]. These proposals are motivated from previous literature [112, 14]. Such heuristic approaches have the downside of requiring careful parameter tuning in order to produce useful particle candidates. We argue that S-PBP produces at least as good particle candidates as heuristic proposal generators without requiring parameter tuning.

Combining DPMP with our S-PBP framework is straightforward. Figure 4.6 illustrates the integration of S-PBP to DPMP in comparison to vanilla DPMP.

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

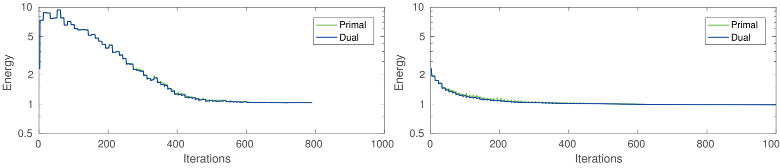


Figure 4.5: Temperature annealing (left) versus diverse particle selection (right) on the image denoising toy example of Sect. 4.1.4.

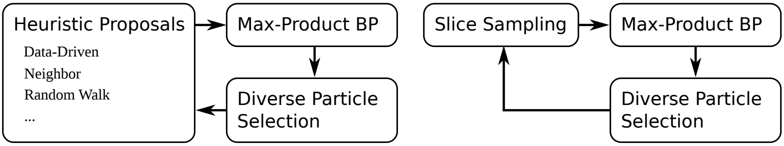


Figure 4.6: Integration of S-PBP in the diverse particle selection framework of [90].

4.2.2 Heuristic Proposals versus Slice-Sampling

In the following the proposed S-PBP approach is compared to black box slice sampling and the heuristic proposal generator in DPMP. The graphical model for this experiment is a chain graph with one hundred nodes only consisting of Gaussian pairwise potentials $\phi_{s,t}(x_s, x_t) \propto \exp\left[-(x_t - x_s - 1)^2/2\sigma^2\right]$ and without unary potentials. The heuristic proposal generator consists of two components: a neighbor-based proposal $q(x_s \mid x_t) = \phi_{s,t}(x_s, x_t)$ for t uniformly sampled from the set of neighbors $\mathcal{N}_s$ and a random walk proposal $q(x_s \mid x_s^{\text{old}}) = \exp\left[-(x_s - x_s^{\text{old}})^2/2\sigma_{\text{rnd}}^2\right]$. Neighbor-proposal and random walk proposals are mixed with probabilities p_{nb} and p_{rw}, respectively, where $p_{\text{nb}} = 1 - p_{\text{rw}}$. Thus, hyperparameter selection is obtained by performing grid search over $p_{\text{rw}} \in \{0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6\}$ and $\sigma_{\text{rnd}} \in \{0.05, 0.1, 0.2, 0.5, 1.0, 2.0, 4.0\}$. As loss function, the area under the curve of the energy with respect to the iteration number is used (omitting the first 20% of iterations for reducing noise induced by random initialization). This loss encourages a fast decreasing energy as well as a low final energy. The estimated hyperparameters are $p_{\text{nb}} = 0.7$, $p_{\text{rw}} = 0.3$, and $\sigma_{\text{rnd}} = 0.2$.

The initial particles $\{x_s^{(i)}\}_{i=1,..,p}$ are drawn randomly in the interval $I = [0, 3 \cdot S]$, where S is the number of nodes. Independent uniform sampling of the initial particles causes high odds of initializing the Markov chain in a poor initial state due to the high probability of leaving a large gap between nearby random samples as illustrated in Fig. 4.7a). This leads in the first iterations of PBP to move towards a local

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

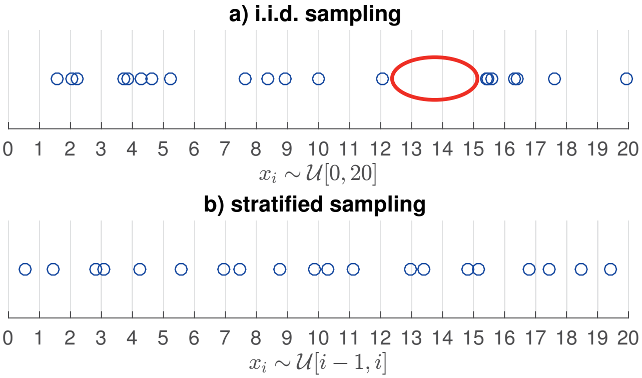


Figure 4.7: a) Independently and identically distributed (i.i.d.) sampling can lead to large gaps as highlighted with the red ellipse. (b) Stratified sampling avoids large gaps by dividing the support region in equal sized subregions and draw exactly 1 sample within each subregion.

optimum with high energy from which it takes a long time to recover. Therefore, instead of i.i.d. sampling we use stratified sampling as follows:

$$x_s^{(i)} \sim \mathfrak{U}([3\cdot(i-1), 3\cdot i) \quad , \quad i = 1,..,p. \tag{4.23}$$

This guarantees that the gap between two particles will never be larger than twice the subregion width, as can be seen in Fig. 4.7b).

Figure 4.8 shows comparison results of DPMP with heuristic proposals, with black-box slice sampling, and with S-PBP proposals, respectively. DPMP was run with $p = 20$ particles per node for all experiments.

Slice sampling and S-PBP produce very low energy MAP estimates after 300 DPMP iterations, while the heuristic proposal method is slower with about 800 DPMP iterations. Black-box slice sampling is the most computationally demanding method, and heuristic sampling is the fastest. S-PBP is a good tradeoff between the other methods as it requires a moderate number of function evaluations while producing high informative samples. While a single S-PBP iteration is slower than heuristic sampling, this was compensated by a faster convergence (i.e. less iterations) as shown in the bottom right of Fig. 4.8.

In the next and final synthetic experiment, the chain graph of above is extended

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

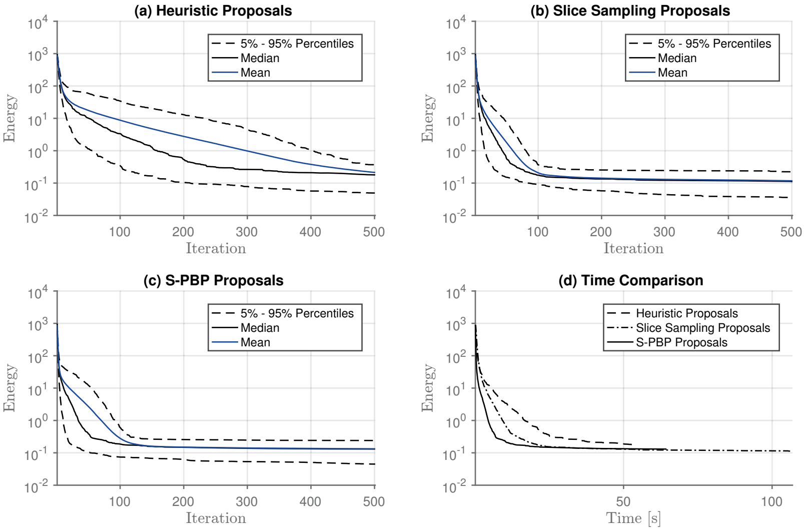


Figure 4.8: Comparison of DPMP with heuristic proposals, slice sampling, and with S-PBP on the chain graph with uninformative unary potentials, respectively. (a)-(c) Evaluation over the number of iterations. Energy mean (blue), median (solid black), 5%, and 95% percentiles (dashed black) are computed over 100 random initialized runs and with 20 particles per node. (d) Comparison of the mean energy of the corresponding methods with respect to run time on an Intel Core i7-3770K 3.50 GHz CPU (4 cores) and 32 GB RAM.

with unary potentials:

$$p(x_1, \ldots, x_S) = \prod_{s=1}^{S} \phi_s(x_s) \prod_{s=1}^{S-1} \phi_{s,s+1}(x_s, x_{s+1}) \tag{4.24}$$

where $\phi_s(x_s)$ are randomly generated Gaussian mixture models and the pairwise potentials are (unnormalized) Gaussian distributions $\phi_{s,s+1}(x_s, x_{s+1}) = \exp[-(x_s - x_{s+1})^2/2\sigma^2]$ with variance $\sigma = 0.1$. The number of nodes is $S = 100$. The number of components (mixtures) per Gaussian mixture model is 10. Due to the Gaussian mixtures, the joint probability $p(x_1, \ldots, x_N)$ is highly nonconvex and hence MAP inference is challenging. The number of particles is $p = 20$. It was observed that DPMP gets stuck at local optima regardless of the number of particles (it was tested with $p = 100$ particles) using either heuristic sampling, slice sampling, or S-PBP.

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

Therefore, to ensure convergence, DPMP is combined with temperature annealing. The annealing scheme of Eq. (4.15) is used with start temperature $T_0 = 200.0$ and end temperature $T_N = 1.0$ where the number of PBP iterations is $N = 200$.

The heuristic sampler consists of three components: sampling from neighborhoods $q_s(x_s) \propto \sum_{t \in \mathcal{N}_s} \phi_{s,t}(x_s, x_t)$, sampling from unary potentials $q_s(x_s) \propto \phi_s(x_s)$, and random walk sampling $q_s(x_s) = \mathcal{N}(x_s \mid {x_s}^{\text{old}}, \sigma^2_{\text{rnd}})$. The hyperparameters for the heuristic sampler are the probabilities for selecting one of the components $p_{\text{nb}}, p_{\text{un}}, p_{\text{rw}}$, respectively, and the random walk variance σ^2_{rnd}. These four parameters were selected using grid search over $p_{\text{un}}, p_{\text{rw}} \in \{0.0, 0.1, 0.2, 0.3, 0.4, 0.5\}$, $p_{\text{nb}} = 1 - p_{\text{un}} - p_{\text{rw}}$, and $\sigma_{\text{rnd}} \in \{0.5, 1.0, 2.0, 4.0, 8.0, 16.0\}$. The optimal parameters for the tested instantiation of the graphical model are $p_{\text{nb}} = 0.7$, $p_{\text{un}} = 0.0$, $p_{\text{rw}} = 0.3$, and $\sigma_{\text{rnd}} = 2.0$.

Figure 4.9 shows the evaluation results. It can be observed that all methods achieve similar energy. Here, heuristic sampling achieves similar performance with respect to the number of iterations. With respect to wall clock time, heuristic sampling is faster than the other methods. Note, however, that heuristic sampling requires careful tuning of the hyperparameters. This is not the case for slice sampling and S-PBP.

4.3 Discussion

In this chapter, a novel proposal generation method for the particle belief propagation (PBP) framework was presented. The proposed approach, slice-sampling particle belief propagation (S-PBP), generates random sample proposals from the (preudo-) max-marginals of the target graphical model distribution using slice sampling. Slice sampling is much less sensitive to hyper-parameters than competing methods. The hyper-parameters in slice sampling only affect the approximation quality and the average number of function evaluations for computing the slice regions. It was shown that by exploiting the message-passing property of PBP, the slice regions can be computed either exactly or approximately for a large variety of factor potential functions using product slice sampling. Thus, dependence on hyper-parameters is eliminated.

It was shown in experiments on synthetic data that the S-PBP proposals outperform Metropolis-Hastings sampling (Sect. 4.1.4) and that they perform on par with heuristic sampling (Sect. 4.2.2). A downside of S-PBP is its relatively high computational effort compared to heuristic sampling. On the other hand, this can often be compensated by the generated high-quality sample proposals leading to low-energy MAP estimates requiring fewer MCMC iterations and PBP iterations than the competing methods.

In the following chapter, the previously developed methods provide the core for inference in both continuous and discrete-continuous graphical models for articulated online object tracking.

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

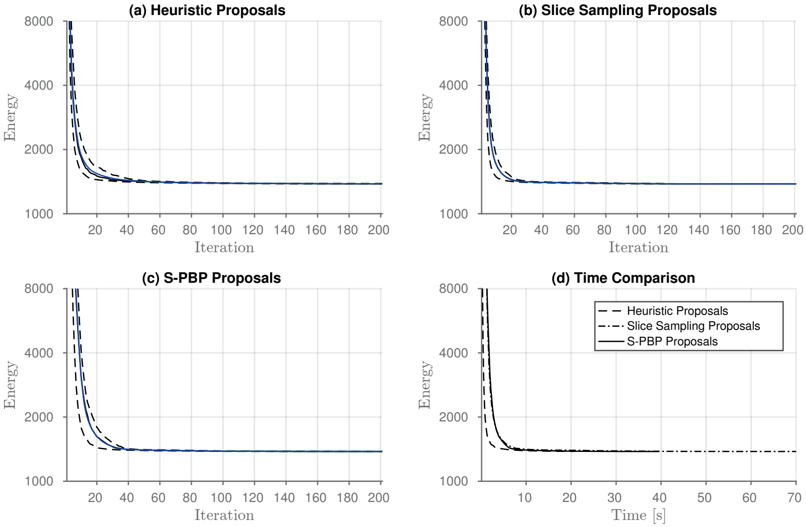


Figure 4.9: Comparison of DPMP with heuristic proposals, with slice sampling, and with S-PBP, respectively, on the chain graph with Gaussian mixture unary potentials. (a)-(c) Evaluation over the number of iterations. Energy mean (blue), median (solid black), 5%, and 95% percentiles (dashed black) are computed over 100 random initializations. (d) Comparison of the mean energy of the corresponding methods with respect to wall clock time.

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

Chapter 5

Tracking

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte
Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

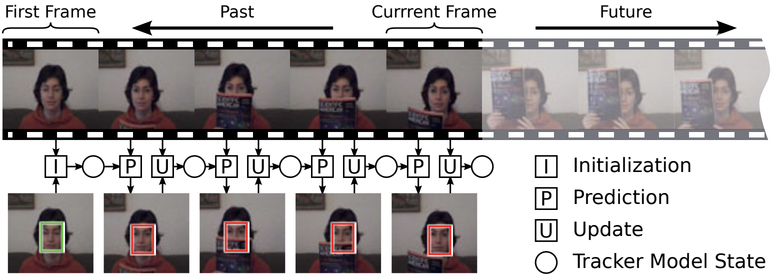


Figure 5.1: Online tracker principle. The tracker is initialized at the first frame with a pose annotation provided by the user (green rectangle). On each new arriving image a *prediction* step estimates the current object pose. The predicted poses are illustrated as red rectangles. The *update* step performs learning using the current (and possibly past) frame and the predicted poses. Looking ahead to future frames is strictly forbidden. This is visually illustrated by fading out of the future frames.

In this chapter, the previously developed inference algorithms (cf. Chapt. 4) are applied to part-based online object tracking. Object tracking is a fundamental problem in computer vision. It is the basis for many high-level tasks such as human-computer interaction, scene understanding, action recognition and many more. *Online* object tracking is a special case of object tracking in which the object class (a human, an animal, a cup, etc.) is a priori unknown. The tracker is initialized using a manual annotation (such as a bounding-box) in the first frame. In a *prediction* step, the object pose is estimated on previously unseen image frames. The *update* step performs a model update, i.e. learning. This step is crucial in order to adapt to appearance changes during tracking. Furthermore, the tracker is not allowed to look at future frames. The frames arrive to the system in a similar fashion to a live feed, e.g., from a web cam. This property is known as *causality.* The workflow for online object tracking is illustrated in Fig. 5.1.

The challenging part of object tracking is when both the target object and the background undergo strong appearance changes or temporal occlusions of the target object occur. Changes of the object appearance can be caused by illumination changes, object deformations (e.g. out-of-plane rotation), occlusions, camera blur, and many more. Object appearance descriptors have to be as invariant as possible against such influences. Although it has been made rapid progress in the field of bounding-box based online tracking [70], state-of-the-art trackers must resort to clever feature descriptors and sophisticated online learning mechanisms. This comes at the price of high computational complexity and poor scalability with respect to

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

the number of target objects to track at the same time. State-of-the-art features are based on deep neural networks [27, 70]. Forward propagation in these networks is known to be computationally high demanding even on (massively) multi-threaded platforms. The results from the very recent VOT challenge 2016 [70] show that all state-of-the-art trackers do not reach real-time processing speed.

It was shown in other works [127, 132] that part-based approaches increase tracker robustness with respect to non-rigid object deformations and partial occlusions when compared to bounding-box trackers. A large amount of changes in appearance can be well compensated by these part-based approaches without resorting to sophisticated feature descriptors. The introduction of parts although comes at the price of increased computational complexity due to the exponential growths on the number of feasible poses. While the search space dimension in bounding-box trackers is relatively small and can be well-compensated by applying efficient convolution operations (such as discrete Fourier transforms), such approaches do not scale well to higher dimensional problems. Efficient inference is achieved in tree-structured models when either the pairwise relationships are Gaussian [37] or a coordinate transform can be applied such that efficient Gaussian convolution can be applied [4]. This, though, does not scale to loopy graphs with more complex pairwise or higher-order relations.

The following sections handle the part-based object tracking problem using the stochastic inference methods proposed in Chapt. 4. Stochastic inference scales well to high-dimensional problems. In Sect. 5.1, a template-based tracker is introduced. The S-PBP method proposed in Sect. 4.1.2 is applied and evaluated against concurring methods. Sect. 5.3 extends the part-based tracking model in several ways. First, the model space is extended to include foreground/background segmentation cues, leading to a discrete-continuous inference problem with complex high-order constraints. Second, the template-based likelihoods are replaced with discriminative models as used in state-of-the-art online trackers. In a third contribution (cf. Sect. 5.3.3), the tracker model is furthermore extended with auxiliary variables modeling the *visibility* of parts.

5.1 Part-Based Template Tracking

The proposed feature tracker uses a pairwise MRF model. The model is separated into two parts: (a) the unary potentials are derived from a feature patch matching model, and (b) the pairwise potentials encode the relative positioning of the features to each other. The label space of the MRF is the space of feature poses including the local central patch position, patch rotation, and scale. The proposed MRF model is as follows:

$$E(\mathbf{x}) = \sum_{s \in \mathcal{V}} \psi_s(x_s) + \alpha \cdot \sum_{s \in \mathcal{V}} \sum_{t \in \mathcal{N}_s} \psi_{s,t}(x_s, x_t), \tag{5.1}$$

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

where the unary potential function

$$\psi_s(x_s) = \chi^2(\mathrm{HOG}_{I_n}(\mathbf{p}_s, \mathbf{o}_s), \mathrm{HOG}_{I^{\mathrm{ref}}}(\mathbf{p}_s^{\mathrm{ref}}, \mathbf{o}_s^{\mathrm{ref}})) \tag{5.2}$$

is the Chi-square distance of histogram of oriented gradients (HOG) features [79] of a patch at position $\mathbf{p}_s \in \mathbb{R}^2$ of the current image I_n and orientation $\mathbf{o}_s \in \mathbb{R}^2$, where $x_s = \{\mathbf{p}_s, \mathbf{o}_s\}$ and a reference image I^{ref} at reference position $\mathbf{p}_s^{\mathrm{ref}}$ and orientation $\mathbf{o}_s^{\mathrm{ref}}$. The orientation vector $\mathbf{o}_s$ encodes two aspects: the feature patch rotation (rotation of $\mathbf{o}_s$, i.e. $\mathrm{atan2}(\mathbf{o}_s)$) and feature patch scale (length of $\mathbf{o}_s$, i.e. $\|\mathbf{o}_s\|_2$). For the tracker to be able to deal with fast moving objects, a *resolution pyramid* approach is applied on the unary potentials. This is done by concatenating HOG descriptors of patches with differing spatial resolution. For each resolution pyramid level (*scale*) the image is downsampled by a factor of 0.5 using bicubic interpolation.

The pairwise potential $\psi_{s,t}(x_s, x_t)$ is as follows:

$$\psi_{s,t}(\,\cdot\,) = \frac{\|\mathbf{p}_t - \mathbf{p}_s - \mathbf{R}_s\mathbf{d}_{st}\|_2^2 + \|\mathbf{p}_s - \mathbf{p}_t - \mathbf{R}_t\mathbf{d}_{ts}\|_2^2}{2 \cdot \|\mathbf{d}_{st}\|_2^2} \tag{5.3}$$

where $\mathbf{d}_{st(ts)} = \mathbf{p}_{t(s)}^{\mathrm{ref}} - \mathbf{p}_{s(t)}^{\mathrm{ref}}$ and $\mathbf{R}_{s(t)} = \begin{bmatrix} o_{x,s(t)} & -o_{y,s(t)} \\ o_{y,s(t)} & o_{x,s(t)} \end{bmatrix}$ is a 2×2 rotation and scale matrix. The proposed pairwise potential function models the surrounding of each feature point as a weak-perspective model and transforms its neighbor points (with respect to the reference frame) according to a similarity transformation (consisting of translation, rotation, and scaling).

The scalar parameter $\alpha > 0$ is a weighting factor determining the *stiffness* of the feature mesh balancing between feature point independence ($\alpha \to 0$; i.e. multi-target tracker) and rigid single object tracking.

5.1.1 Inference

To infer the most probable pose $\mathbf{x} = \min E(\mathbf{x})$ in each frame, the S-PBP approach as introduced in Chapt. 4 is applied. Inference is done for each frame independently, but the particle set from the previous frame is used as a starting point for the current frame. In order to increase tracker robustness, a *particle resampling* step is applied where for each frame the initial set of particles is drawn with replacement from the set of particles $\{x_s^{(i)}\}_{i=1,\ldots,p}$ from the previous frame with probability $\hat{\mu}_s(x_s^{(i)})$.

S-PBP requires the computation of slice intervals $A_{\psi_s}(u)$ and $A_{\psi_{s,t}}^{x_t}(u)$. Since $\psi_{s,t}$ is given as an analytic function (cf. (5.3)), an analytic derivation of the slice intervals can be obtained by using a symbolic solver. The MuPAD® computer algebra system is used for fully automatic symbolic slice region computation. An analytic form of the data-driven unary potentials is not available and thus a reasonable approximation for the corresponding slice intervals has to be found. We chose to set the slice interval $A_{\psi_s}(u)$ to the whole image space for $\mathbf{p}_s$, i.e. $\mathbf{p}_s \in [1,W] \times [1,H]$, where

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

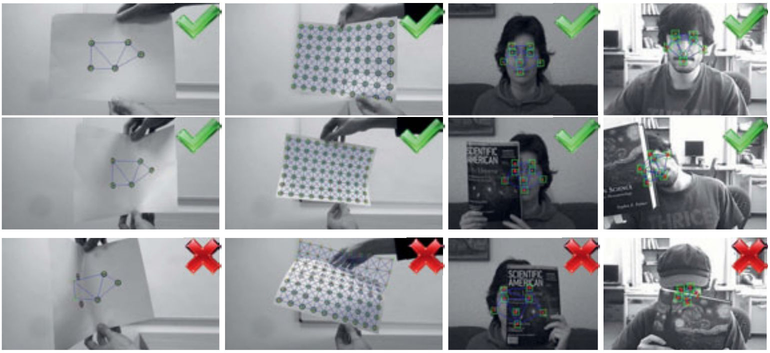


Figure 5.2: Datasets and tracking results for our proposed method: PAPER1, PAPER2, FACEOCC1, FACEOCC2 (from left to right). First two rows: successful tracking; third row: tracking failure cases.

W and H are the image width and height respectively, and to restrict $\mathbf{o}_s$ to $\mathbf{o}_s \in [-10{,}10] \times [-10{,}10]$. This way it is ensured that the sampling space is large enough. On the other hand, particles sampled outside the true slice regions are discarded by resorting to the rejection sampling mechanism in Alg. 6.

The S-PBP approach is compared to the MH-PBP approach. In order to provide a fair comparison the the proposal distribution has to designed carefully such that good MCMC chain mixing is ensured. We propose to use a 4D Gaussian distribution with a covariance matrix Σ combined with a suitable coordinate transformation. The label space can be divided into two parts, the feature position $\mathbf{p}_s \in \mathbb{R}^2$ and orthogonal feature transformation $\mathbf{o}_s \in \mathbb{R}^2$. The proposal distribution for $\mathbf{p}_s$ is $p(\mathbf{p}_s^{\langle m \rangle} \mid \mathbf{p}_s^{\langle m-1 \rangle}) = \mathcal{N}(\mathbf{p}_s^{\langle m \rangle} \mid \mathbf{p}_s^{\langle m-1 \rangle}, \mathbf{I}_{2\times 2} \cdot \sigma_{xy})$, where $\mathcal{N}(x \mid \mu, \Sigma)$ is a Gaussian distribution over x with mean μ and covariance Σ. $\mathbf{I}_{2\times 2}$ is the 2×2 identity matrix. The vector $\mathbf{o}_s$ is sampled analogously, but in the polar coordinate system with covariance matrix $\Sigma_{\text{polar}} = [\sigma_r^2, 0; 0, \sigma_\phi^2]$, where σ_r^2 is the variance for the radius and σ_ϕ^2 the variance for the angle. The proposal distribution depends on three hyperparameters σ_{xy}, σ_r, and σ_ϕ, which are tuned using grid search.

5.1.2 Experiments

Test sequences

The following challenging test sequences are used for evaluation: PAPER1, PAPER2, FACEOCC1, and FACEOCC2. The self-made PAPER1 and PAPER2 sequences were cho-

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

sen to challenge the methods on a fast moving deformable object under major scale changes and deformations. The sequences have a spatial resolution of 960 px × 540 px and consist of 563 and 726 frames respectively. The captured object (paper) is textured with patches of similar appearance and shape. The similar appearing features were chosen to stress the relational structure of our tracker model. Thus the only way to distinguish the features is by considering the relative position of the feature patches to each other. The PAPER1 sequence consists of five feature patches with a carefully chosen position pattern which allows unique identification of the features by only having knowledge about the relative distances of the features to each other. The PAPER2 sequence is more challenging since the number of features is increased to 70 and the features are arranged in a grid structure allowing local relational ambiguities. The FACEOCC1 and FACEOCC2 sequences from [5, 30] are designed for evaluating object trackers under major occlusions. The sequences have a spatial resolution of 352 px × 288 px (FACEOCC1) and 320 px × 240 px (FACEOCC2) and both consist of 888 frames each. While the FACEOCC1 sequence has only slow object movements, but showing substantial occlusions, the FACEOCC2 sequence challenges with fast movements, illumination changes, object rotation and substantial occlusions. The sequences and tracking results are shown in Fig. 5.2.

Parameter selection

Parameter selection can be split into two parts. The first part consists in MRF model parameter selection. Since the proposed model is relatively robust to changes in α, this parameter is set in an ad-hoc fashion for each sequence as follows: $\alpha = 20$ for PAPER1 and PAPER2 and $\alpha = 50$ for FACEOCC1 and FACEOCC2. For the HOG features we set the smallest scale pyramid resolution to 50 px × 50 px. This leads to 3 scales for FACEOCC1 and FACEOCC2 and 4 scales for PAPER1 and PAPER2.

The second part is parameter selection for the PBP framework. The number of PBP iterations is set to $N = 20$ and the number of particles to $p = 10$ for each node. With this setting both algorithms (MH-PBP and S-PBP) converge well. Since the overall *sampling* behavior of the proposed method is evaluated rather than the *belief propagation* convergence behavior, selecting these parameters should be uncritical.

Evaluation metrics

The distance $\varepsilon_{\text{track}}$ between the estimated feature position and the groundtruth (manually labeled) position is used as a quality measure. Two metrics are derived from this measure: The root-mean-square deviation (RMSD) and a quantile box-plot (10%, 25%, 50%, 75%, and 90% quantiles). While the first metric is very sensitive to outliers, the second metric provides more information about the overall error distribution.

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

5.1.3 Discussion

The evaluation results comparing S-PBP with MH-PBP using different MCMC iterations are shown in Fig. 5.3. For MH-PBP, the MH sampling parameters $\{\sigma_{xy}, \sigma_r, \sigma_\phi\}$ are chosen (from the set $\{0.1,0.2,0.5,1.0,2.0,5.0\} \times \{0.01,0.02,0.05,0.10,0.20,0.50\} \times \{0.01,0.02,0.05,0.10,0.20,0.50\}$) such that the RMSD is minimized. Note that for S-PBP such parameter tuning is not necessary. We have evaluated the tracking performance for different MCMC iterations $M = 2$ to 5. The box plots in Fig. 5.3 show that S-PBP outperforms or performs equally well as MH-PBP for all tested sequences except for sequence PAPER2 with only 2 (and 3) MCMC iterations where both methods fail. This is mainly due to a much higher overall sampling noise of the MH-PBP method compared to S-PBP. We observed that the sampling noise of S-PBP is much less than with MH-PBP at feature positions with high confidence (i.e. high belief). On the other hand the sampling noise of S-PBP increases for uncertain feature positions. The RMSD in sequence PAPER2 and FACEOCC1 is higher for S-PBP than for MH-PBP due to temporal tracking failures. These tracking failures are caused by strong local deformations or by occlusions of many feature points. Typical tracking failures are depicted in the bottom row of Fig. 5.2. It can be observed in such cases that S-PBP leads to much higher tracking error than MH-PBP due to broader particle sampling in uncertain feature positions.

Figure 5.4 shows an evaluation of MH-PBP under differing (non-optimal) sampling parameters. To this end, we vary each of the three sampling parameters individually and let the other two parameters stay fixed at their optimal values. Note that the estimation error varies highly, where very high values (usually $> 15\,\text{px}$) indicate a tracking failure. In order to visualize both the performance differences for near-optimal parameters and tracking failures, the error values below and above the 15 px mark are shown with a differing vertical axis scaling. In Fig. 5.4, only a comparison for PAPER1 and FACEOCC1 is shown. The other two sequences perform similarly. It can be observed that the tracking performance of MH-PBP strongly depends on careful parameter selection. The parameter σ_{xy} has the highest impact on the tracking performance and the optimal parameter value varies strongly between sequences ($\sigma_{xy} = 5$ for PAPER1 and $\sigma_{xy} = 0.5$ for FACEOCC1). Selecting σ_{xy} is a compromise between allowing fast object motions and reducing overall localization noise. Selecting σ_r and σ_ϕ has analogous effects on changes in object scaling and rotation. This way one has to incorporate *prior knowledge* about the object motion in order to obtain good tracking results using MH-PBP.

The computational complexity for MH-PBP is $\mathcal{O}(NSpM\ (1 + Vp))$ and for S-PBP is $\mathcal{O}(NSpM(3 + 2Vp))$ given the number of PBP iterations N, nodes S, particles p, MCMC iterations M and the average number of neighbors per node V. This indicates a doubling of computation time of S-PBP compared to MH-PBP which is due to the overhead introduced for computing the interval bounds A. A look at the CPU times using fixed parameters for both algorithms ($M = 5$, $p = 10$, $N = 20$)

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

verifies this finding: FACEOCC: 0.69 s/frame (S-PBP) vs. 0.33 s/frame (MH-PBP) ; PAPER2: 7.43 s/frame vs. 3.66 s/frame. Nevertheless it was shown that S-PBP needs significant less MCMC iterations than MH-PBP such that the computational overhead can be typically well compensated.

5.2 Object Tracking Demonstrator

An online part-based face tracker demonstrator was developed on top of the relational feature tracker as proposed in the previous Sect. 5.1. The tracker runs in real-time and is robust against partial occlusions, lighting and appearance changes. Since the tracker's goal is to track faces and to get rid of manual part initialization, the template-based likelihoods (cf. Eq. (5.2)) are replaced in favor of discriminative HOG features which are trained using a linear SVM [114].

The tracker was embedded in a live demonstrator for controlling a game via visual input from a webcam, as shown in Fig. 5.5. The system was exposed on the CeBIT 2015 [73].

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

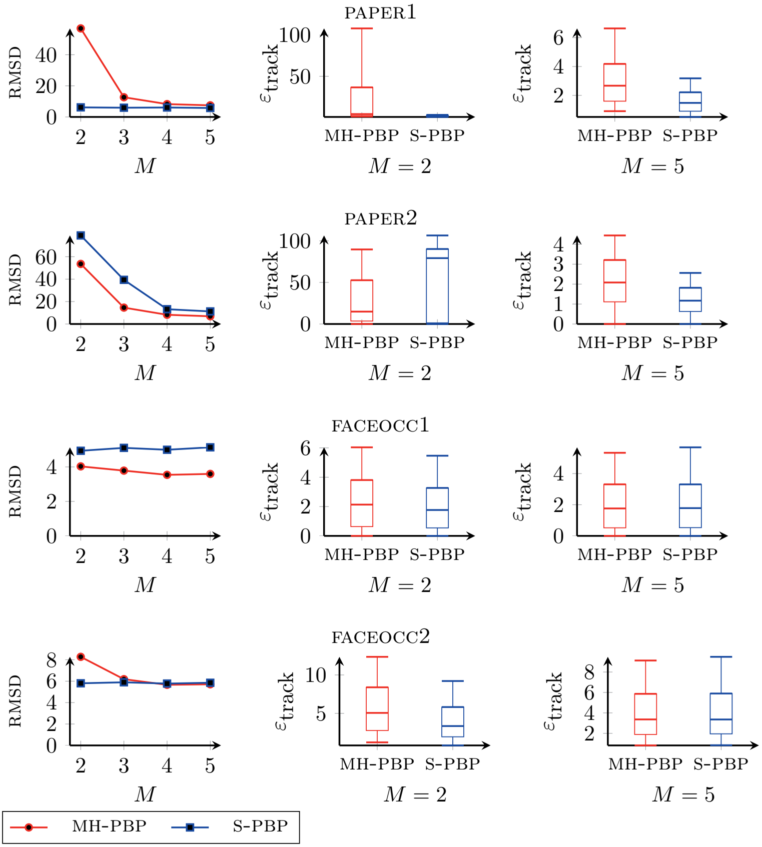


Figure 5.3: Relational feature tracker evaluation results showing the overal RMSD (for MCMC iterations from 2 to 5) and box plots over the error distance to groundtruth for selected MCMC iterations.

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

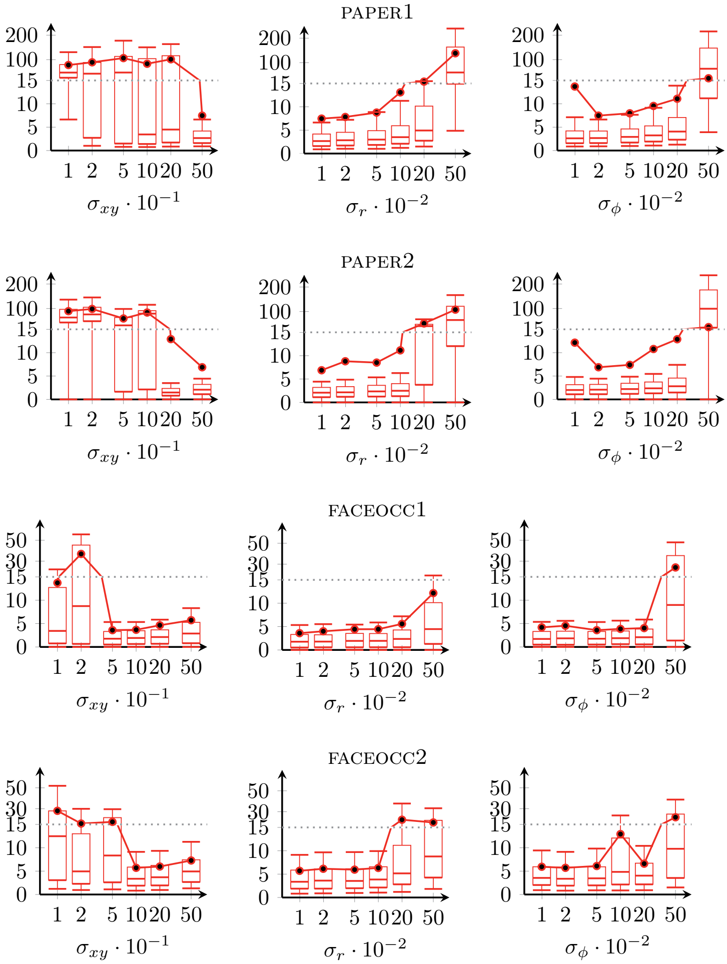


Figure 5.4: Optimal parameter evaluation for MH-PBP method (with $M = 5$). The vertical axis shows the error distance to groundtruth in px. Note that the vertical axis is stretched for error values lower than 15 px in order to better visualize performance differences.

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

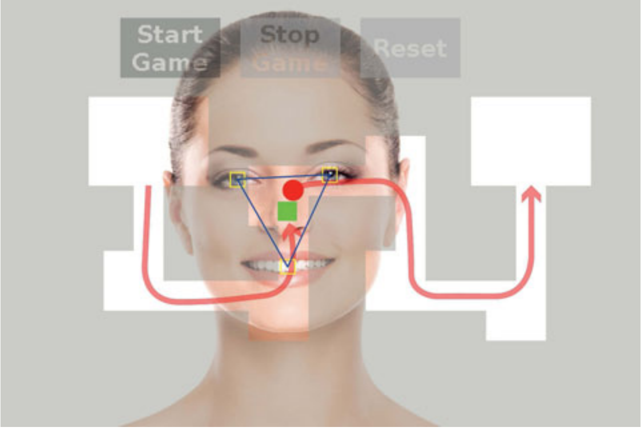


Figure 5.5: Realtime part-based face tracking demonstrator: Schematic overview of the game flow.

5.3 Joint Tracking and Segmentation

Bounding-box initialized visual object tracking has made rapid progress in the last couple of years [71, 124]. Especially the use of online learning methods [44, 56, 132] and recently the incorporation of part-based models [132, 127] and image segmentation [119, 74] have lead to significant improvements in visual online tracking.

Another rapidly evolving area is part-based pose estimation using offline trained discriminative models such as DPMs [36, 125, 32]. DPMs are favorably applied to human pose estimation. Hereby, the human body is decomposed into semantically meaningful parts. The relations between the parts are modeled using weak joints such as Gaussian-type pairwise probability density functions. Unfortunately, an expensive offline-training phase is necessary, requiring a huge database of annotated full poses.

The goal in this section is to merge both approaches and propose a part-based online object tracking framework providing detailed articulated pose while keeping the initialization and training effort low. The proposed model is automatically initialized from a single template image and a foreground pixel mask, as shown in Fig. 5.6. Automatic object initialization is a major issue due to (self-)occlusions occurring in the template image. No prior semantic or structural information is available which define, for example, a skeleton. To tackle this problem, the DPM is extended with (a) part visibility cues to model (self-)occlusions and (b) a visibility-

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

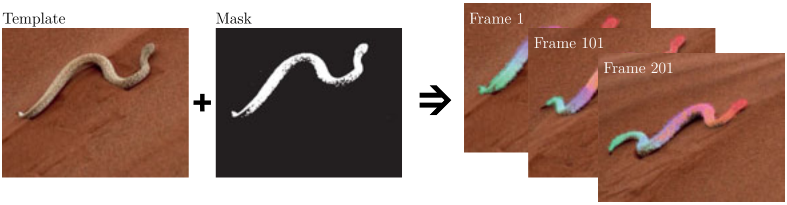


Figure 5.6: Fine-grained non-rigid object tracking with strong segmentation cues. The DPM is generic (non-discriminative) and solely initialized from the template image (left). The parts are color-coded and overlaid to the input frames (best viewed in color).

adapted spring-force model which allows to adaptively *switch off* part dependencies.

Suitable priors are crucial for a stable visibility estimation on the parts level. To this end, *patch masks* are introduced which define the shape of the object parts. The visibility weighted patch masks are combined and forced to consistency with a foreground/background image segmentation through a novel *global consistency prior*. The basic principle of the global consistency prior is depicted in Fig. 5.7 (yellow box).

In the online tracking domain, state-of-the-art approaches [44, 56, 132, 127] lead to high quality bounding box estimates. The trackers in [44, 56] parametrize the target object pose as a single rigid patch. Recent approaches [132, 127] combine online tracking with part-based models.

Part-based models such as DPMs [36] are well known approaches for pose estimation and detection of highly articulated objects such as humans. They have been originally introduced for object detection. They use a relatively simple parametrization with only 2D positions and a spring force related model to connect the parts in a tree structure. It was shown by [125] that the expressiveness of DPMs can be significantly improved by increasing the number of patches combined with a “switched spring model”. A multi-layer arrangement of patches can further improve pose estimation accuracy [32]. Due to the inherent tree-structure, these methods still suffer from the well known double-counting problem.

The methods [119, 74] deal with (self-)occlusions and double-counting by fusing pixel-wise image segmentation with DPM. Ladicky et al. [74] assign image pixels to a part-specific label and enforce consistency between image segmentation and DPM by “switching off” the body parts which do not correspond to the image segmentation labeling. We follow their approach and introduce *visibility* variables which modify the influence of the appearance and pairwise DPM terms. The work in [119] enforces consistency between DPMs and a FG/BG segmentation by introducing pairwise po-

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

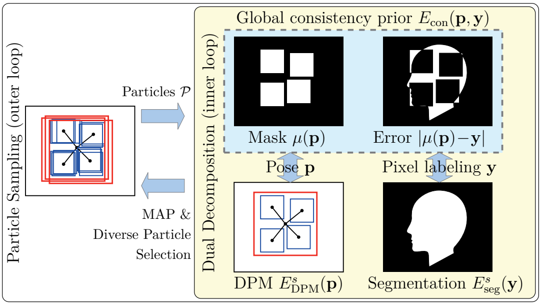

Particle Sampling (outer loop)
Particles $\mathcal{P}$
MAP & Diverse Particle Selection
Dual Decomposition (inner loop)
Global consistency prior $E_{\text{con}}(\mathbf{p},\mathbf{y})$
Mask $\mu(\mathbf{p})$
Error $|\mu(\mathbf{p})-\mathbf{y}|$
Pose $\mathbf{p}$
Pixel labeling $\mathbf{y}$
DPM $E^s_{\text{DPM}}(\mathbf{p})$
Segmentation $E^s_{\text{seg}}(\mathbf{y})$

Figure 5.7: Schematic overview of our joint pose estimation and image segmentation using particle sampling (left) and MAP inference via dual decomposition (yellow area). Our proposed global consistency prior (blue area, cf. Eq. (5.7)) enforces consistency between pose estimation and image segmentation.

tentials loosely coupling each image pixel with each part. For a tight coupling they introduce non-convex high-order residual terms. Since optimizing their full objective function is intractable, they use a relaxed DD [65, 66] approach for inference. They propose a lower bound approximation of the dual energy by ignoring the higher-order terms. This leads to an over-estimate of the primal-dual gap and hence important properties of DD are lost. We adopt the DD approach but propose an approximation of the higher-order constraints to enable efficient inference of the full objective function.

The contributions in this section are as follows: A novel yet simple formulation of a global consistency prior for joint part-based object tracking and image segmentation is proposed. It is shown how to solve the global consistency constraints efficiently using methods based on dual decomposition. This model is further extended to predict the visibility of each DPM part. The additional visibility variables are coupled with the DPM pose prior. This can lead to increased robustness towards suboptimal model initialization and topology changes during tracking of highly articulated subjects. The joint tracking and segmentation framework is evaluated on the VOT 2015 and VOT 2016 benchmark [71, 70] (cf. Sect. 2.1) and on manually annotated sequences. The proposed joint model consistently outperforms the baseline trackers.

First, the basic framework for joint post estimation and image segmentation is introduced in Sect. 5.3.1. Following that, this framework is extended with latent visibility variables in Sect. 5.3.3. An evaluation of both frameworks is conducted in Sect. 5.3.4, followed by a discussion in Sect. 5.3.5.

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

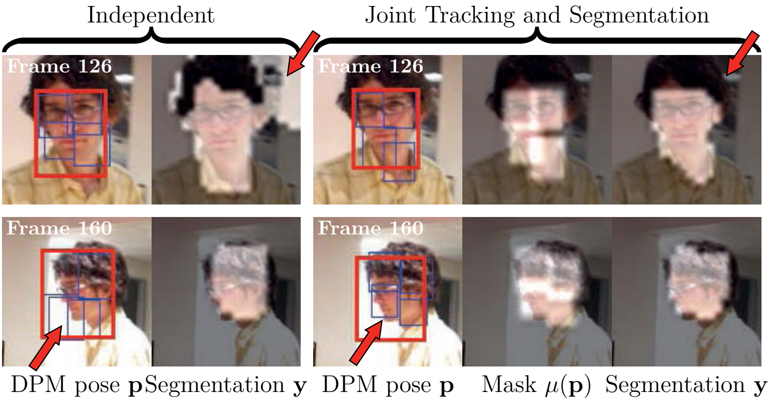


Figure 5.8: Online tracking framework: Comparison of joint tracking and segmentation (right) and independent tracking and segmentation (left) on two frames from the *david* sequence. Top: The DPM guides the image segmentation. Bottom: The image segmentation guides the part placement of the DPM. Differences are highlighted by red arrows.

5.3.1 Joint Pose Estimation and Segmentation

Object tracking over a sequence of frames $s = 1,\ldots,S$ can be cast as a frame-by-frame pose estimation problem [127, 132], where a pose $\mathbf{p}^s$ and pixel-wise labeling $\mathbf{y}^s$ is estimated for each frame $s \geq 2$. Frame $s = 1$ is the reference frame. The joint pose estimation and image segmentation model is based on the following problem formulation

$$\{\mathbf{p}^s, \mathbf{y}^s\} = \underset{\mathbf{p}\in\mathcal{X}^{\text{pose}},\mathbf{y}}{\operatorname{argmin}}\ E^s_{\text{DPM}}(\mathbf{p}) + E^s_{\text{seg}}(\mathbf{y}) + E_{\text{con}}(\mathbf{p},\mathbf{y}). \tag{5.4}$$

The first term corresponds to a deformable part model. The second term is a pixel-wise image segmentation energy, labeling every pixel as either foreground or background. The third term is a consistency constraint enforcing consistency between image segmentation and pose estimation. A comparison of a joint optimization to an independent handling of the two complementary cues E_{DPM} and E_{seg} is visualized in Fig. 5.8. The three terms in Eq. (5.4) are introduced in detail in the following paragraphs.

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

Deformable Parts Model

The proposed framework is built upon the well known deformable parts model (DPM) [36], where the target object is decomposed into a set of parts. The object parts are indexed over a node set $\mathcal{V}$. Each part $i \in \mathcal{V}$ has its own pose configuration p_i which includes the 2D-position and scale. The parts are related to each other by local spring forces, which encode pairwise spatial context information. The set of these pairwise relations (the edge set) is denoted by $\mathcal{E}$. The energy function over the corresponding graphical model $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is as follows

$$E^s_{\mathrm{DPM}}(\mathbf{p}) = \sum_{i \in \mathcal{V}} f^s_i(p_i) + \sum_{(i,i') \in \mathcal{E}} f^s_{ii'}(p_i, p_{i'}), \tag{5.5}$$

where $\mathbf{p} = (p_1, \dots, p_{|\mathcal{V}|}) \in \mathcal{X}^{\mathrm{pose}}$, f^s_i are unary potentials encoding an appearance model (e.g. obtained by a score map from a discriminative classifier), and the binary potentials $f^s_{ii'}(p_i,p_{i'})$ encode directed spring forces.

Efficient MAP inference is done via max-product belief propagation over a finite set of candidate poses $p_i \in \mathcal{P}_i = \{p^1_i, ..., p^{L_i}_i\}$. A commonly followed approach in online tracking is to densely sample pose candidates over the whole search space by applying grid search with a spacing of one pixel. Inference over such a large number of candidates is not tractable in our case since the patch masks, introduced later, then will easily exceed RAM limitations. Hence an iterative stochastic sampling approach is applied instead based on DPMP (cf. Chap. 4). New particle candidates can be either generated from a proposal distribution $q_i(p_i)$ or with S-PBP.

Image Segmentation

The image segmentation model consists of a pixel-wise appearance term and a pairwise contrast-sensitive Potts-model between neighboring pixels. The corresponding energy function is as follows

$$E_{\mathrm{seg}}(\mathbf{y}) = \sum_{j \in \mathcal{W}} g_j(y_j) + \sum_{(j,j') \in \mathcal{F}} d_{jj'} \mathbb{I}[y_j \neq y_{j'}], \tag{5.6}$$

where $\mathbf{y} = (y_1, \dots, y_N) \in \mathcal{X}^{\mathrm{seg}} = \{0,1\}^N$, $N = |\mathcal{W}|$, $\mathcal{W}$ is the set of pixels, $\mathcal{F}$ is the set of edges (a 4-neighborhood), $d_{jj'}$ is a contrast-sensitive weight similar to [19] and $\mathbb{I}[\,\cdot\,]$ is the indicator function. Normalized *Lab* color histograms with $8 \times 8 \times 8$ bins are used as appearance model $g_j(y_j)$. The energy term $E_{\mathrm{seg}}(\mathbf{y})$ is submodular and hence can be efficiently optimized with graph cuts [6, 18, 17, 64].

Global Consistency Prior

The consistency prior should enforce that every pixel which is covered by (at least) one part of the DPM is labeled as foreground. All other pixels are labeled as background. On the other hand, if some pixels prefer to be foreground due to the image

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

segmentation appearance model, at least one part should cover those pixels. Consequently, pixels preferring to be background should not be covered by any part.

A *masking* function $\mu_j : \mathcal{X}^{\text{pose}} \to [0,1]$ for each pixel j is introduced which encodes the area overlap of the DPM parts with the j-th image pixel, i.e. $\mu_j(\mathbf{p}) = 1$ if a part covers pixel j and $\mu_j(\mathbf{p}) = 0$ if no part covers pixel j. For pixels at the boundary of a part, $\mu_j(\mathbf{p})$ can be something between 0 and 1 (e.g. via anti-aliasing). Stacking all functions μ_j together leads to the vectorized form $\mu : \mathcal{X}^{\text{pose}} \to [0,1]^N$. Note that this is not restricted to hard mask boundaries. In fact, it can be advantageous to define *soft* masks where the part boundaries appear blurred. This accounts better for uncertainties in the part shape modeling.

The consistency constraint energy can be formulated as

$$E_{\text{con}}(\mathbf{p}, \mathbf{y}) = \alpha \|\mathbf{y} - \mu(\mathbf{p})\|_2^2 = \alpha \sum_{j \in \mathcal{W}} (y_j - \mu_j(\mathbf{p}))^2 \tag{5.7}$$

which is a pixel-wise quadratic soft constraint enforcing $\mu_j(\mathbf{p}) \approx 1$, if Pixel j is foreground, and $\mu_j(\mathbf{p}) \approx 0$, if Pixel j is background.

This constraint is in general very hard to optimize due to its non-convexity with respect to $\mathbf{p}$. The mask function is approximated as follows, leading to tractable optimization of the consistency constraint. The mask function is decomposed into independent parts analogously to the DPM formulation:

$$\mu_j(\mathbf{p}) \approx \textstyle\sum_{i \in \mathcal{V}} \mu_{ij}(p_i). \tag{5.8}$$

Furthermore, the pose $\mathbf{p}$ is restricted to a discrete set of pose candidates $p_i \in \mathcal{P}_i = \{p_i^1, ..., p_i^{L_i}\} \subset \mathcal{X}_i^{\text{pose}}$, i.e. $p_i = p_i^{l_i}$ for some $l_i \in \{1, ..., L_i\}$. This way, the pose vector $\mathbf{p}$ can be recoded as an indicator variable vector $\mathbf{x} = [\mathbf{x}_1; ...; \mathbf{x}_{|\mathcal{V}|}]$ with $\mathbf{x}_i = \mathbf{e}^{L_i}(l_i)$. The vector $\mathbf{e}^L(l)$ is an L-dimensional column-vector where the l-th entry is one and all other entries are zero. These constraints are encoded in the constraint set $\mathcal{C}$, i.e. $\mathbf{x} \in \mathcal{C}$. The mask function $\mu(\,\cdot\,)$ can thus be encoded as a linear function

$$\mu_j(\mathbf{p}) = \textstyle\sum_{i \in \mathcal{V}} \boldsymbol{\mu}_{ij} \cdot \mathbf{x}_i = \boldsymbol{\mu}_j^{\mathsf{T}} \mathbf{x}, \tag{5.9}$$

where $\boldsymbol{\mu} = [\boldsymbol{\mu}_1, ..., \boldsymbol{\mu}_N]^{\mathsf{T}}$ is a (sparse) matrix. The joint energy term with indicator variables reads as

$$\min_{\mathbf{x} \in \mathcal{C}, \mathbf{y} \in \mathcal{D}} E(\mathbf{x}, \mathbf{y}) = E_{\text{DPM}}(\mathbf{x}) + E_{\text{seg}}(\mathbf{y}) + \alpha \|\mathbf{y} - \boldsymbol{\mu}\mathbf{x}\|_2^2. \tag{5.10}$$

Optimization of this non-convex function is still non-trivial. In the following section, a dual decomposition approach is proposed to solve problem (5.10).

5.3.2 Dual Decomposition

The problem (5.10) is solved with dual decomposition (cf. Sect. 3.3.3). The proposed decomposition of problem (5.10) consists of three sub-problems: the DPM, the segmentation, and the global consistency constraint. The decomposition is graphically depicted in Fig. 5.7.

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

The decomposition is derived as follows. A second set of variables $\mathbf{x}', \mathbf{y}'$ is introduced. The equality constraints $\mathbf{x} = \mathbf{x}'$ and $\mathbf{y} = \mathbf{y}'$ enforce them to be exact copies of $\mathbf{x}$ and $\mathbf{y}$:

$$\begin{aligned} \min_{\substack{\mathbf{x}\in\mathcal{C},\mathbf{y}'\in\mathcal{D}\\ \mathbf{x}',\mathbf{y}}} \quad & E_{\mathrm{DPM}}(\mathbf{x}) + E_{\mathrm{seg}}(\mathbf{y}) + \alpha\|\mathbf{y}' - \boldsymbol{\mu}\mathbf{x}'\|_2^2 \\ & \text{s.t.} \quad \mathbf{x} = \mathbf{x}', \quad \mathbf{y} = \mathbf{y}'. \end{aligned} \tag{5.11}$$

Through Lagrangian relaxation, the hard equality constraints are relaxed, leading to the following Lagrangian dual formulation:

$$\begin{aligned} \max_{\lambda_1,\lambda_2} \min_{\substack{\mathbf{x}\in\mathcal{C},\mathbf{y}'\in\mathcal{D}\\ \mathbf{x}',\mathbf{y}}} \; & E_{\mathrm{DPM}}(\mathbf{x}) + E_{\mathrm{seg}}(\mathbf{y}) + \alpha\|\mathbf{y}' - \boldsymbol{\mu}\mathbf{x}'\|_2^2 \\ & + \lambda_1^{\mathsf{T}}(\mathbf{x} - \mathbf{x}') + \lambda_2^{\mathsf{T}}(\mathbf{y} - \mathbf{y}') \end{aligned} \tag{5.12}$$

With some rearrangement, the Lagrangian dual can be decomposed into the following sub-problems (also referred to as *slave* problems):

$$g_1(\lambda_1) = \min_{\mathbf{x}\in\mathcal{C}} E_{\mathrm{DPM}}(\mathbf{x}) + \lambda_1^{\mathsf{T}}\mathbf{x} \tag{5.13}$$

$$g_2(\lambda_2) = \min_{\mathbf{y}\in\mathcal{D}} E_{\mathrm{seg}}(\mathbf{y}) + \lambda_2^{\mathsf{T}}\mathbf{y} \tag{5.14}$$

$$g_3(\lambda_1, \lambda_2) = \min_{\mathbf{x}',\mathbf{y}'} \alpha\|\mathbf{y}' - \boldsymbol{\mu}\mathbf{x}'\|_2^2 - \lambda_1^{\mathsf{T}}\mathbf{x}' - \lambda_2^{\mathsf{T}}\mathbf{y}' \tag{5.15}$$

with the corresponding *master* problem

$$\max_{\lambda_1,\lambda_2} g_1(\lambda_1) + g_2(\lambda_2) + g_3(\lambda_1, \lambda_2). \tag{5.16}$$

Sub-problems 1 and 2 take the same form as Eqs. (5.5) and (5.6), respectively. Hence, exactly the same methods can be used for solving the corresponding sub-problems. Sub-problem 3 is an unconstrained quadratic program and hence leads to solving a (sparse) linear equation system. However, sub-problem 3 is unfortunately not well defined since $\boldsymbol{\mu}^{\mathsf{T}}\boldsymbol{\mu}$ is singular almost always. Hence, regularization is introduced:

$$\begin{aligned} g_3(\lambda_1, \lambda_2) = \min_{\mathbf{x}',\mathbf{y}'} \quad & \alpha\|\mathbf{y}' - \boldsymbol{\mu}\mathbf{x}'\|_2^2 - \lambda_1^{\mathsf{T}}\mathbf{x}' - \lambda_2^{\mathsf{T}}\mathbf{y}' \\ & + \beta_1\|\mathbf{x}'\|_2^2 + \beta_2\|\mathbf{y}' - 0.5\|_2^2, \end{aligned} \tag{5.17}$$

where the regularization parameters β_1 and β_2 are chosen sufficiently small. The regularization of $\mathbf{y}'$ towards 0.5 has the rationale of not introducing a bias neither towards foreground ($y_j = 1$) nor towards background ($y_j = 0$). It is easy to show that the solution of the joint problem in Eq. (5.4) stays the same for all β_1,$\beta_2 \geq 0$.

The master problem is concave and non-smooth and hence can be solved using a subgradient method:

$$\lambda^{k+1} = \lambda^k + \eta_k \nabla g(\lambda^k) \tag{5.18}$$

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

where λ^k is the concatenation of λ_1 and λ_2 at the k-th iteration. The subgradient is

$$\nabla g(\lambda^k) = [\mathbf{x}^k - \mathbf{x}'^k;\ \mathbf{y}^k - \mathbf{y}'^k] \tag{5.19}$$

where $\mathbf{x}^k$, $\mathbf{y}^k$, $\mathbf{x}'^k$, and $\mathbf{y}'^k$ are the minimizer of the respective sub-problems $g_1(\lambda_1^k)$, $g_2(\lambda_2^k)$, and $g_3(\lambda_1^k, \lambda_2^k)$. The step-size is $\eta_k = \gamma \cdot \frac{\text{bestPrimal}-\text{Dual}}{\|\nabla g(\lambda^k)\|^2}$ for a positive scalar γ, following the heuristic step size rule in [65, 66].

Finally, a primal solution $(\mathbf{x}^*, \mathbf{y}^*)$ is constructed from the partial solutions $\bar{\mathbf{x}}, \bar{\mathbf{x}}', \bar{\mathbf{y}}$, and $\bar{\mathbf{y}}'$ with $\mathbf{x}^* = \bar{\mathbf{x}}$ and $\mathbf{y}^* = \operatorname{argmin}_{\mathbf{y}} E_{\text{seg}}(\mathbf{y}) + E_{\text{con}}(\mathbf{x}^*, \mathbf{y})$. That is, the solution of the DPM subproblem is used and from this, the optimal segmentation given the DPM pose is computed. This can, again, be efficiently computed with graph cuts. The best primal solution (wrt. Eq. (5.10)) over all DD iterations is returned as the final MAP estimate.

The final inference approach is a double-loop algorithm (cf. Fig. 5.7). In each outer loop iteration a new particle set is sampled using DPMP. The inner loop handles the dual decomposition updates over the sampled pose candidates as described above.

Since the inner loop problem (5.11) is expected to be quite similar from iteration to iteration, we warm start DD by initializing the Lagrangians $\lambda_{1,2}$ accordingly from the previous DD run. Initializing λ_2 is trivial, since the state space $\mathcal{X}^{\text{seg}}$ for $\mathbf{y}$ is static. Warm-starting the Lagrangians λ_1 is more involved as the particles $\mathcal{P} \subset \mathcal{X}^{\text{pose}}$ for $\mathbf{p}$ change after each outer loop iteration. An interpolation strategy is used with $\lambda_1^{\text{new}} = \text{interp}_{\mathcal{P}^{\text{old}} \rightarrow \mathcal{P}^{\text{new}}}(\lambda_1^{\text{old}})$ to map the state space defined by the particle set of the previous iteration $\mathcal{P}^{\text{old}}$ to the state space $\mathcal{P}^{\text{new}}$ of the current iteration. A simple nearest-neighbor interpolation strategy is applied.

Segmentation with Shape Prior

To show the efficacy of the proposed consistency prior, a preliminary image segmentation experiment is conducted as shown in Fig. 5.9. The focus is set on testing the influence of the consistency prior in presence of a weak DPM and image segmentation model. The DPM model only consists of a single part and hence no pairwise connections are involved. The DPM appearance model is kept completely uninformative. Thus, pose estimation is solely guided through the consistency prior. The part has a square shape as shown in the left image of Fig. 5.9 and can vary in its x,y-position and square side length s. This shape information is encoded in the mask function μ. Since the DPM is completely uninformative, particles are sampled from a proposal function $q_i(p_i)$. The x,y-position is sampled from a Gaussian random walk proposal $\mathcal{N}([x,y]^\mathsf{T} \mid [x,y]^\mathsf{T}_{\text{old}}, \sigma^2_{\text{rw}})$ and the scale s is sampled from a Gamma distribution $\Gamma(s \mid k,\theta)$ with shape k and scale θ chosen such that the mode $(k-1)\theta = s_{\text{old}}$ and the variance $k\theta^2 = \sigma^2_{\text{scl}}$. The parameters are set to $\sigma_{\text{rw}} = \sigma_{\text{scl}} = 0.2$.

The input image is corrupted with severe Gaussian and salt/pepper noise which makes the image segmentation model quite uncertain in its decision. As can be

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

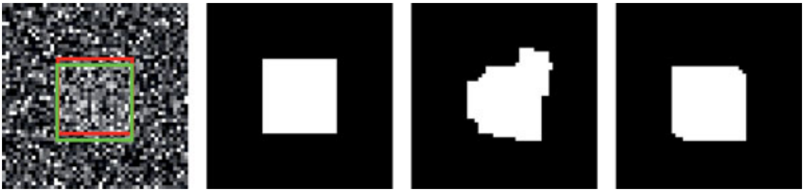


Figure 5.9: Image segmentation with shape prior. Left: Input image corrupted with noise, groundtruth pose (red square), reconstructed pose (green square). Middle left: Groundtruth segmentation. Middle right: Graph cut segmentation. Right: Our approach.

seen in the right image of Fig. 5.9, our consistency prior leads to an improved segmentation compared to a segmentation using $E_{\text{seg}}(\mathbf{y})$ only. Furthermore, the correct pose $\mathbf{p}$ could be reconstructed (green rectangle).

5.3.3 Visibility Estimation

The model presented in Sect. 5.3.1 implicitly assumes that all DPM parts are visible in every frame. This is a reasonably well assumption for short-time object tracking. In fact, this assumption is exploited in state-of-the-art online tracking frameworks where the *best* pose candidate in the current frame is selected as a new training sample candidate, such as is the case with the popular Struck tracker [44]. These approaches can well compensate for temporal distractions (e.g. occlusions) by using robust training methods. Thus, explicit visibility estimation often cannot bring performance improvements.

On the other hand, when strong image segmentation cues are available, visibility estimation can significantly *improve* tracking performance. In the following, it is assumed that the foreground/background image segmentation cues are sufficiently reliable. Auxiliary variables $\mathbf{v} = (v_1, \ldots, v_{|\mathcal{V}|})$ are introduced to the DPM model, where $v_i \in [0,1]$ with $v_i = 0$ if part i is invisible and $v_i = 1$ if part i is fully visible. The DPM in section 5.3.1 is thus modified as follows:

$$E_{\text{VDPM}}(\mathbf{p}, \mathbf{v}) = \sum_{i \in \mathcal{V}} v_i f_i(p_i) + \sum_{(i,i') \in \mathcal{E}} v_i v_{i'} f_{ii'}(p_i, p_{i'}). \tag{5.20}$$

Note that the visibility estimates not only influence the part appearance term, but also the connections between parts. The rationale behind this is that it is often difficult to provide appropriate pose priors in presence of topology changes or severe deformations (cf. Figs. 5.10, 5.12, and 5.13). Especially in the case of online tracking, the amount of available training data is not enough to provide well-trained,

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

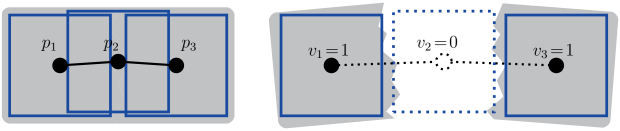


Figure 5.10: Visibility-aware DPM: Through introducing part weights v_i, the DPM parts (blue rectangles) are allowed to *vanish* (dotted rectangle) and its corresponding spring forces (black lines) are weakened accordingly (dotted lines). Thus, the *ripping apart* of an object can be modeled by the visibility-aware deformable parts model (VDPM) if the underlying image segmentation (gray area) supports this.

complex pose priors which can handle strong deformations and arbitrary non-rigid articulations. On the other hand, developing generative pose priors is a non-trivial task [1]. The introduction of latent variables can bridge the gap between the generative and discriminative approaches. Here, we use the variables v_i to *switch on/off* the spring forces $f_{ii'}(p_i, p_{i'})$. Likewise, the mask function $\mu(\mathbf{p})$ is augmented with $\mathbf{v}$ to $\mu(\mathbf{p}, \mathbf{v})$ such that $\mu_{ij}(p_i, v_i) = v_i \cdot \mu_{ij}(p_i)$. This accounts for part transparency.

Multi-Layer Model

Note that a *flat* model such as is commonly used in human pose estimation methods [36] tends to rip apart into two or more completely independent objects. This is due the tree structure of the DPM, where per definition, every part is connected to other parts over a unique path (cf. Fig. 5.11, top). We counteract this behavior by introducing long-range dependencies through a multi-layer DPM as shown in Fig. 5.11, bottom.

Inference

The pair $(\mathbf{p}, \mathbf{v})$ can, again, be encoded with indicator variables $\mathbf{x}$, with the difference that the non-zero entries in $\mathbf{x}$ are not 1 but in $[0, 1]$. To be precise: $\mathbf{x}_i = v_i \mathbf{e}^{L_i}(l_i)$. This constraint is encoded in the constraint set $\mathcal{C}' \supset \mathcal{C}$. Since the image segmentation is assumed to be almost perfect, the energy term $E_{\text{seg}}(\mathbf{y})$ can be dropped. To nevertheless account for a data-driven penalty of deviations of the image segmentation to the mask $\boldsymbol{\mu}\mathbf{x}$, we introduce a latent variable vector $\mathbf{w} \in [0,1]^N$ and a penalty term $\theta_{\text{seg}}^{\mathsf{T}} \mathbf{w}$. The modified energy function is as follows

$$\min_{\mathbf{x} \in \mathcal{C}', \mathbf{w} \in [0,1]^N} E_{\text{VDPM}}(\mathbf{x}) + \theta_{\text{seg}}^{\mathsf{T}} \mathbf{w} \quad \text{s.t.} \quad \mathbf{w} = \boldsymbol{\mu}\mathbf{x}. \tag{5.21}$$

Unfortunately, the Lagrangian relaxation as used in Sect. 5.3.1 is not tight. This can be seen by considering the subproblem $\min_{\mathbf{w} \in [0,1]^N} (\theta_{\text{seg}} + \lambda_2)^{\mathsf{T}} \mathbf{w}$. Here, the solution

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

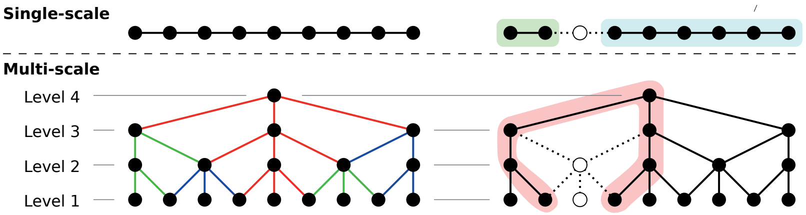


Figure 5.11: Behaviour of single-scale model (top) vs. multi-scale model (bottom) in case when some parts are not visible (non-filled circles). In the single-scale model, the disappearance of one part leads to a separation into two independent sub-graphs (green and blue areas). In the multi-scale model, the connection is retained through the higher scale levels (red area).

is always integer, since the objective function is linear. Hence, no solution with $w_j \in (0,1)$ could be found although $\boldsymbol{\mu}_j^\top \mathbf{x}$ could span anything between $[0,1]$ which contradicts with the equality constraint $\mathbf{w} = \boldsymbol{\mu}\mathbf{x}$ in problem (5.21). To resolve this issue, we resort to an *augmented* Lagrangian relaxation approach, thus leading to the non-convex alternating direction method of multipliers (ADMM) (cf. [16, 81]). The augmented Lagrangian is of the following form

$$\begin{aligned}\mathcal{L}(\mathbf{x},\mathbf{w},\mathbf{x}',\mathbf{w}',\mathbf{u}) = & E_{\mathrm{VDPM}}(\mathbf{x}) + \theta_{\mathrm{seg}}^\top \mathbf{w} + \tfrac{\rho}{2}\|\mathbf{x}-\mathbf{x}'+\mathbf{u}_1\|_2^2 + \\ & \tfrac{\rho}{2}\|\mathbf{w}-\mathbf{w}'+\mathbf{u}_2\|_2^2 \quad \text{s.t.} \quad \mathbf{w}' = \boldsymbol{\mu}\mathbf{x}' \end{aligned} \tag{5.22}$$

This form is in general difficult to optimize. The ADMM approach is to alternate between (i) minimizing over $(\mathbf{x},\mathbf{w})$, and holding $\mathbf{x}',\mathbf{w}',\mathbf{u}$ fix, (ii) minimizing over $(\mathbf{x}',\mathbf{w}')$, and holding $\mathbf{x},\mathbf{w},\mathbf{u}$ fix, and (iii) a gradient ascent step over $\mathbf{u}$ with $\mathbf{u}^{k+1} = \mathbf{u}^k + \rho([\mathbf{x};\mathbf{w}] - [\mathbf{x}';\mathbf{w}'])$. The $(\mathbf{x}',\mathbf{w}')$-update corresponds to solving (5.15) plus a quadratic term. Performing the $(\mathbf{x}',\mathbf{w}')$-update is essentially the same as solving the DD subproblem g_3 (cf. Eq. (5.15)), since both problems are (unconstrained) quadratic programs. The $(\mathbf{x},\mathbf{w})$-update involves the optimization of $E_{\mathrm{VDPM}}(\mathbf{p},\mathbf{v})$. A joint optimization is not tractable and we resort to alternating between optimization over $\mathbf{p}$ and $\mathbf{v}$.

An extension to loopy DPMs, as required by the multi-scale model, is straight forward. The loopy DPM is decomposed into a set of trees $\mathcal{T}$ with $E_{\mathrm{DPM}}(\mathbf{x}) = \sum_{t\in\mathcal{T}} E_{\mathrm{DPM}}^t(\mathbf{x})$ and each tree has its own set of pose variables $\mathbf{x}^t$. The set of consensus constraints (cf. Eq. (5.11)) is extended to $\mathbf{x}^t = \mathbf{x}'$, accordingly. See [65] for further details. The used tree decomposition is indicated in Fig. 5.11 (bottom left) via color-coding of the lines.

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

Connection to Flexible Mixture-of-Parts

The proposed model can be interpreted as an extension to the flexible mixtures-of-parts (FMP) model [125]. FMP is a switched-spring model, extending the DPM with hidden variables $t_k \in \{1,...,T_k\}$ which activate t_k-th spring from a set of T_k springs per part:

$$E_{\mathrm{FMP}}(\mathbf{p},\mathbf{t}) = \sum_{k \in \mathcal{V}'} f_k^{t_k}(p_k) + \sum_{(k,k') \in \mathcal{E}'} f_{kk'}^{t_k,t_{k'}}(p_k, p_{k'}). \qquad (5.23)$$

A connection of this approach to ours can be seen by setting $f_i(p_i) = f_k^{t_k}(p_k)$, $f_{ii'}(p_i,p_{i'}) = f_{kk'}^{t_k,t_{k'}}(p_k, p_{k'})$ for a bijective mapping $\imath : (k,t_k) \mapsto i$ for all $k = 1, ..., |\mathcal{V}|$, $t_k = 1, ..., T_k$, and the additional constraint

$$\textstyle\sum_{t_k=1}^{T_k} v_{\imath(k,t_k)} = 1 \quad , \quad v_i \in \{0,1\} \qquad (5.24)$$

stating that *exactly one* spring has to be chosen for each part. By dropping the constraint (5.24), the model is relaxed such that *at most one* part (and its connected springs) can be active at a certain position. At the same time, it is ensured by the global consistency constraint E_{con} that at least one part and corresponding springs become activated in areas where the foreground cues are strong enough. In contrast to [125], the proposed model is not prone to the well-known double counting phenomenon (i.e. the activation of multiple parts at a single location), since an activation of multiple parts at the same position would lead to $\boldsymbol{\mu}\mathbf{x} \gg 1$, which is prevented by the E_{con} constraint.

5.3.4 Experiments

The proposed approach is evaluated on two application cases. First, bounding-box initialized short-term online tracking. Here, the objects are rather compact in shape such that the object pose can be well described using a simple bounding-box. The main challenges here are strong appearance changes and partial occlusions during tracking. The second case is non-rigid object tracking with severe articulation. Here, bounding-box estimation is not sufficient for describing the object pose. It will be shown that in both cases, the incorporation of the global consistency prior leads to improved pose estimation results.

Visual Object Tracking Benchmark

The proposed tracker in Sect. 5.3.1 is evaluated on the VOT 2015 benchmark dataset [71] comprising 60 sequences of varying length, image size and tracking difficulty. As appearance model, we use the Struck tracker [44]. This tracker provides an support vector machine (SVM) score map $S(p)$ which can be converted to a probability map

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

through Platt scaling. Each part $i \in \mathcal{V}$ of the DPM has its own tracker with unary terms

$$f_i(p_i) = -\alpha_{\text{score}} \cdot \log\left[\tfrac{1}{1+\exp\{-S_i(p_i)\}}\right]. \tag{5.25}$$

The part structure consists of five parts in two scale layers (cf. Fig. 5.7), similar to related work [127]. The initial part structure is derived from the initial groundtruth bounding box such that part 1 has the same size and position as the bounding box. Parts 2 to 5 are half the size of part 1 and are connected to part 1 via directed spring forces of type

$$f_{ii'}(p_i, p_{i'}) = \frac{1}{2\sigma^2}\left[\|\frac{\mathbf{d}_{ii'}}{s_i} - \bar{\mathbf{d}}_{ii'}\|_2^2 + \frac{1}{\bar{s}_{ii'}^2}\|\frac{\mathbf{d}_{ii'}}{s_{i'}} - \bar{\mathbf{d}}_{ii'}\|_2^2\right] \tag{5.26}$$

where $\mathbf{d}_{ii'} = [x_i, y_i]^\intercal - [x_{i'}, y_{i'}]^\intercal$ is the displacement of the patch centers of patch i to patch i', and $\bar{\mathbf{d}}_{ii'}$ is the initial displacement in the reference frame. The pose of the i-th part consists of the 2D patch center position (x_i, y_i) and a positive scale factor s_i w.r.t. the initial patch size, i.e. $p_i = (x_i, y_i, s_i)$. The positive factor $\bar{s}_{ii'}$ is the ratio of the patch sizes of the i-th to the i'-th patch in the reference frame.

A neighbor-based proposal distribution $q_i(p_i) \propto \exp[-f_{ii'}(p_i, p_{i'})]$ is used for particle sampling, where i' is chosen randomly from the set of neighbors of part i. Note that no explicit motion prior is applied. Instead, the particles (part pose candidates produced by the DPMP framework) of the last frame are used as initial particles for the new frame. If some particles are previously discarded by the DPMP framework, the contingent is filled up by sample proposals randomly selected from a uniform 7×7 grid around the last bounding box position with a grid spacing of 10 px. Additionally, a temperature annealing scheme is applied with $\text{temp}^{k+1} = 0.96 \cdot \text{temp}^k$, and $\text{temp}^1 = 1.0$ in order to produce less noisy pose estimates.

Parameter tuning of our tracker was done on selected sequences from the online tracking benchmark [124]. These sequences are not part of the VOT 2015 sequences, thus avoiding any overfitting in the final benchmark. Parameters are set empirically to $\alpha = 1.0, \beta_1 = \beta_2 = 0.1, \gamma = 0.1, \alpha_{\text{score}} = 100.0, \sigma = 0.5$.

The proposed approach was tested against the baseline Struck tracker [44], a DPM version of Struck (which is essentially applying our approach with $\alpha = 0.0$), and the single-part Struck tracker with the consistency constraint (i.e. $|\mathcal{V}| = 1$). The proposed approach will be further abbreviated with *Struck-DPM-Seg*, and the other approaches with *Struck*, *Struck-DPM*, and *Struck-Seg*, respectively. The evaluation results are summarized in Tab. 5.1. The joint tracking and segmentation approaches outperform both the *Struck* and *Struck-DPM* approaches without consistency constraints. Note that the proposed framework is not limited to the Struck tracker, but can be applied to any tracker which provides a likelihood map.

In a second experiment, the proposed approach is tested with the state-of-the-art online tracker *CCOT* [27]. We observed that the neighbor-based proposal generation

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

Table 5.1: VOT 2015 benchmark tracking results for different tracker configurations of *Struck* showing the raw accuracy, the average number of failures, and the expected overlap corresponding to [71]. Red is first and blue is second ranking. ↑/↓ indicates higher/lower value is better, respectively.

Tracker	Exp. Overlap↑	Accuracy↑	Failures↓
Struck	0.1960	0.46	2.42
Struck-DPM	0.1839	0.47	2.19
Struck-Seg	**0.2152**	0.47	2.28
Struck-DPM-Seg	0.2020	**0.48**	**1.92**

Table 5.2: VOT 2016 benchmark tracking results for different tracker configurations of *CCOT* showing the raw accuracy, the average number of failures, and the expected overlap corresponding to [70]. Red is first and blue is second ranking. ↑/↓ indicates higher/lower value is better, respectively.

Tracker	Exp. Overlap↑	Accuracy↑	Failures↓
CCOT	0.2807	**0.51**	**1.03**
CCOT-DPM	0.2364	0.48	1.32
CCOT-Seg	**0.2857**	**0.51**	1.07
CCOT-DPM-Seg	0.2676	0.49	1.10

leads to very slow convergence when the state space is extended to scale estimation, instead of only the bounding-box center position. Therefore, the heuristic proposals are replaced in favor of the slice sampling proposal as introduced in Sect. 4.2. The evaluation results on the VOT 2016 benchmark are shown in Tab. 5.2. In overall, joint segmentation and tracking again increases the expected overlap compared to the baseline methods. Multi-part tracking did not improve performance on the *CCOT* tracker.

Fine-scale Non-rigid Object Tracking

In this section, the visibility-aware multi-part tracking framework is evaluated. The VDPM approach is applied on surveillance sequences of babys and on a sequence of a snake. The tracked subjects underly severe articulations, where bounding-box trackers are not suitable. Due to the low number of training data, discriminative DPM approaches with pre-trained, complex pose priors [125, 74, 32] are not applicable. In order to capture the whole body articulations, a fine-grained patch decomposition

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

is desirable. Patch sizes of $10 \times 10\,\mathrm{px}$ for the *Baby*-sequence and $5 \times 5\,\mathrm{px}$ for the *Sidewinder*-sequence with 3 scale levels for both sequences are used. The part likelihoods consist of HOGs and color histograms in red, green, and blue (RGB)-space. All histograms are concatenated to a single feature vector and the χ^2 distance of the current feature to a template feature is used as appearance potential $f_i(p_i)$.

For the FG/BG segmentation, a skin detector is applied on the *Baby*-sequence and k-means clustering on the *Sidewinder*-sequence. It is assumed that the image segmentation is robust enough to allow inference of part visibility.

Groundtruth The videos are manually annotated with skeletal points as shown in the first row in Figs. 5.12–5.15 (blue lines). From this rather sparse representation, the groundtruth patch positions which densely cover the target object are derived (cf. second row in Figs. 5.12–5.15). These patches are generated from the sparse control points as follows. Let $q = (q_1, ..., q_K)$ be the set of control points of the skeletal annotation and $\mathcal{E}_{\text{skelet}}$ be the bone (edge) structure with $e = (i,j) \in \mathcal{E}_{\text{skelet}}$, iff q_i and q_j are connected by a bone (an edge). A Gaussian weight function $g_e(p) = \exp(-0.5(p - \mu_e)^\mathsf{T}\Sigma_e(p - \mu_e))$ over patch positions p is estimated for each bone, indicating the proximity of patch p to bone e. The mean μ_e is chosen as lying exactly in the middle of the bone, i.e. $\mu_e = \frac{q_i+q_j}{2}$. The covariance Σ_e is chosen such that the joints q_i and q_j lie exactly on the focal points of the ellipse $(p - \mu_e)^\mathsf{T}\Sigma_e(p - \mu_e) = 1$ and the ratio $r = \frac{a}{b}$ of the half axes a and b of the ellipse is fixed to some value $r < 1$. The ratio r determines how “skinny” the object appears. We chose $r = 0.5$ in our experiments. The Gaussians are depicted in Figs. 5.12–5.15 as green ellipses.

Evaluation The full VDPM approach is compared to the baseline DPM approach without visibility and segmentation cues. Furthermore, a weaker VDPM model is constructed where the visibility variables only influence the unary potentials but not the pairwise terms. This modification is denoted as visibility-aware deformable parts model without edge (VDPM-e). It corresponds to setting high confidence to the DPM pose prior.

Figures 5.14a and 5.15a show a percentage of correct parts (PCP) evaluation of the different approaches with different spring force parameters σ. VDPM clearly outperforms the baseline DPM and the weaker VDPM-e approach. For VDPM and DPM, the best PCP is reached with $\sigma = 10$ in the *Sidewinder* sequence. Therefore, this value is kept for further experiments. Figues 5.12 and 5.13 rows 3–5 show qualitative results on both sequences. It can be observed that the DPM approach (last row) tends to drift away in presence of high articulations or fails to correctly estimate the motions of the limbs. VDPM-e performs better, but faces problems with self-occlusions during initialization (Fig. 5.13 row 4) and is suspect to local minima (Fig. 5.12 row 4).

Convergence statistics are shown in Figs. 5.14b and 5.15b, where E_{feas} is the pro-

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

jection error of $\mathbf{x}'$ on $\mathcal{C}'$. It can be observed that the non-convex ADMM fails to converge with DPM. The multi-layer DPM consists of many loops and thus finding a consensus of all DPM trees is hampered. VDPM tends to perform better, probably because contradictory local estimates lead to a weak consensus and therefore (temporally) reduces the visibility of such patches. This, in turn, leads to higher consensus. In later iterations, the visibility can rise again with consensus on all trees.

5.3.5 Summary

A framework for bounding-box initialized online tracking and fine-grained articulated object tracking was presented. A joint DPM and image segmentation model is proposed on the basis of a novel, yet simple global consistency prior. Two scenarios are shown in which the joint model improves image segmentation and/or object tracking performance. In the first scenario, the segmentation cues and pose cues are fused to improve object tracking performance. The proposed approach outperforms the baseline *Struck* tracker on the VOT 2015 benchmark. The second scenario handles articulated tracking with strong segmentation cues and weak pose cues. It was shown that extending the standard DPM with auxiliary visibility variables improves tracking performance.

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte
Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

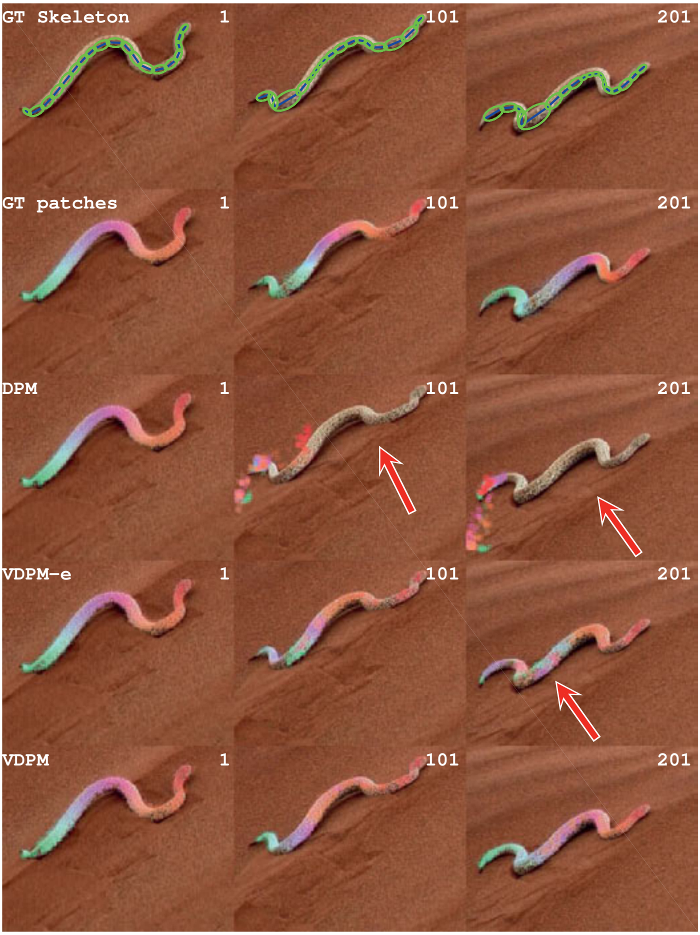


Figure 5.12: Qualitative evaluation results for the SIDEWINDER sequence with 208 frames, 501 × 401 px resolution, and 775 patches. The patch size in the finest scale is 5 × 5 px. The five rows are: Groundtruth skeleton annotation, groundtruth patch positions, DPM, and our proposed approaches VDPM-e and VDPM (best viewed in color). The first column is the reference frame. Red arrows indicate tracking failures.

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

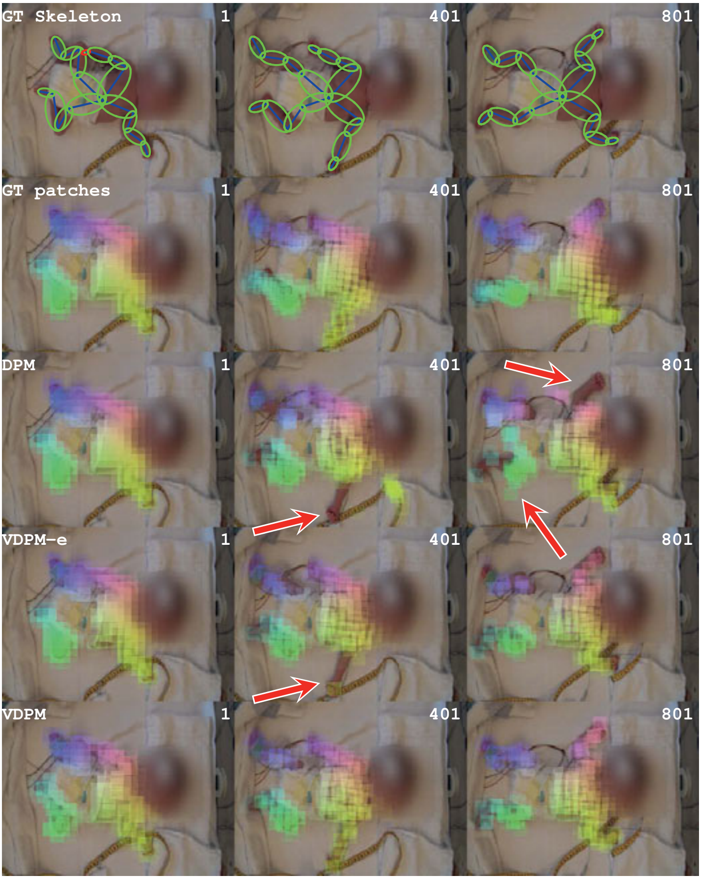


Figure 5.13: Results for the Baby sequence with 828 frames, 360×270 px resolution, and 337 patches. Patch size in the finest scale is 10×10 px. The right leg occludes the hand during model initialization (red circle in the left top image). Red arrows indicate failures.

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

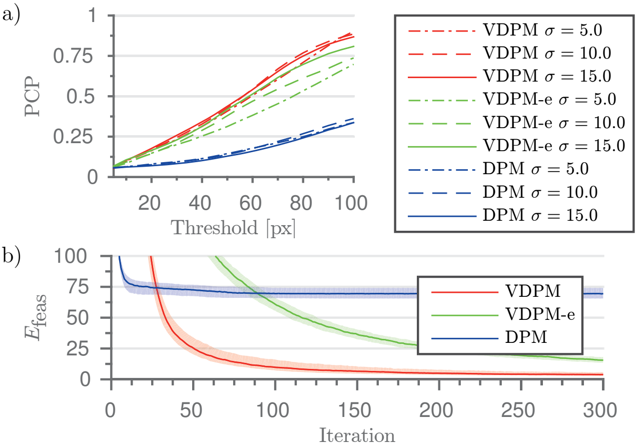


Figure 5.14: a) Quantitative evaluation and b) convergence evaluation results for the SIDEWINDER sequence.

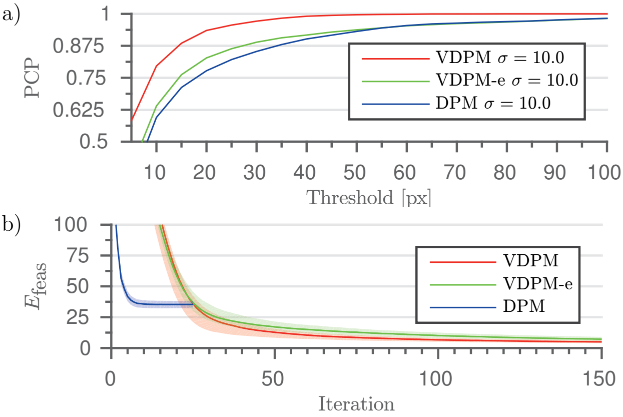


Figure 5.15: a) Quantitative evaluation and b) convergence evaluation results for the BABY sequence.

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

Chapter 6

Conclusions

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

Throughout this thesis, the topic of stochastic inference in continuous-state probabilistic graphical models (PGMs) has been adressed. The guiding application of the developed inference methods is situated in the field of part-based online tracking of articulated objects.

The first contribution of this thesis is the extension of the state-of-the-art on stochastic inference in PGMs towards novel methods for generating more efficient particle proposals. This topic is of particular interest, since the particle proposal generation process is a cruicial part in obtaining good maximum a posteriori (MAP) estimates in high-dimensional space. Previous approaches heavily rely on heuristic proposals which are application-dependent and require carefully tuning of hyper parameters.

The main result of this thesis is a flexible sampling framework for inference in continuous-labeled PGMs based on (product) slice sampling. Our proposed slice-sampling particle belief propagation (S-PBP) algorithm can be either used as a black-box sampler, or a white-box sampler. Black-box sampling is relatively slow and agnostic to the underlying graphical model structure but allows for rapid prototyping. The novel white-box sampling approach has the advantage of fast sample generation and does not dependend on tuning parameters. This is achieved by exploiting the message-passing structure of state-of-the-art inference methods and the incorporation of prior knowledge about the factor potentials of the PGM.

We proposed to use slice sampling in favor of heuristic proposal generators or Metropolis-Hastings based samplers. The incorporation of product slice sampling in the diverse max-product framework leads to a significant speed-up of the MAP inference process. We proposed to exploit the message-passing structure of max-product algorithms to replace the slice region approximation process by exact slice region computation. This approach is applicable as long as exact slice region computation is feasible for the factor graph potentials. In cases where this is not feasible (for instance in high-dimensional data-driven likelihood potentials), the framework reverts to approximate slice region computation. This way, the proposed framework is applicable in both generative and discriminative PGMs.

It was shown that slice sampling leads to less-correlated samples than Metropolis-Hastings while at the same time being more robust towards the choice of hyper-parameters. A downside of our approach is that a single S-PBP iteration is in comparison to other proposal generators much slower. On the other hand, this behaviour is well compensated by faster convergence of S-PBP. S-PBP requires significantly less iterations than all other tested approaches while at the same time producing MAP estimates with lower energy.

The second contribution of this thesis is to combine the field of articulated pose tracing and online tracking. Hereby, the previously developed inference framework provides the core of our proposed part-based online tracking approaches. We started by developing a generative part-based object tracker which is initialized from the first frame of a video or online video stream (for instance from a webcam) by man-

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte
Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

ually annotating the object parts and relations. The proposed tracker is robust against partial occlusions and flexible enough to adapt to deformable surfaces. A live demonstrator application in form of a visual tracker controlled game illustrates the real-time capability of our proposed approach.

The final contribution is an automatic initialized online tracking framework for articulated objects. The basic idea for automatic model initialization is to use an over-segmentation of the target object into a regular grid of small part patches. The main problem of part-based online tracking is tracker drift. We showed that by introducing image segmentation cues and a novel global consistency prior which connects image segmentation and object tracking in a high-order constraint, our framework is able to perform tracking of highly articulated objects with significantly less tracker drift than standard deformable parts model (DPM) based approaches. The resulting objective function is highly non-convex and consists of factor potentials of high order. We proposed a novel combination of particle max-product and dual decomposition for solving the challenging MAP inference task. Experiments show that the proposed joint tracking and segmentation framework improves both part tracking, as well as image segmentation accuracy. On the one hand, the image segmentation cues guide the part tracking. On the other hand, the DPM in conjunction with the proposed global consistency prior works as a shape prior for the image segmentation.

6.1 Future Work

Our proposed S-PBP algorithm is very generic and applicable to a large variety of other PGMs. Applications in which particle-based stochastic inference approaches have already been successfully applied are optical flow estimation and protein folding [90], human pose estimation [104, 90], and self-localization in sensor networks [51].

The particle-based stochastic inference approach still depends on a number of parameter such as the number of particles or the Gibbs temperature and other parameters related to the message passing approaches (number of iterations, damping factor, edge probabilities). All these parameters still require careful tuning in order to guarantee that the algorithm converges to the true MAP. As pointed out before [51], the required number of particles highly depend on the complexity of the factor potentials. Complexity estimation is still an unresolved problem. Furthermore, the Gibbs temperature controls how peaky the joint distribution is and indirectly influences the diversity of particles produced by DPMP. Preliminary experiments show that a low Gibbs temperature (a distribution with sharp peaks and broad, flat valleys) leads to clustering of particles towards a single mode. Future research should focus on developing rules, heuristics, or guidelines for choosing the hyperparameters optimally with respect to balance between computational efficiency and approximation error.

Further research requires the extension of slice sampling in the particle belief

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

propagation framework to the higher-dimensional case. The proposed approach works well in factor graphs with pairwise potentials, but sampling in higher-order potentials is currently only supported by applying Gibbs sampling. Gibbs sampling is suboptimal when the high-dimensional distribution is skewed.

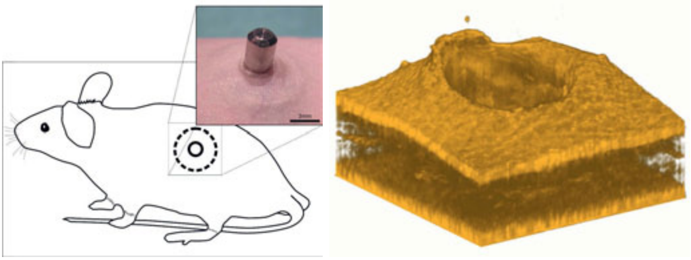
The proposed non-rigid online tracker is currently only implemented in MATLAB® with only partial multi-threaded CPU support. Therefore, the implementation is very slow. As the proposed approach is solely based on distributed inference methods such as dual decomposition and message passing, we expect that for instance a massive parallel GPU implementation would speed up the inference process drastically.

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte
Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

Appendix

Appendix A

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte
Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.



(a) (b)

Figure A.1: (a) Location and photo of a percutaneous implant, and (b) OCT dense 3D scan volume rendering (percutaneous pin is not visible).

A.1 Optical Coherence Tomography

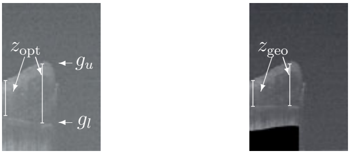
This section summarizes the contributions of the author on OCT motion compensation and image undistortion. This work was published in [84, 85, 29].

OCT is a non invasive imaging modality used for taking optical biopsies of layered tissue structures such as the epidermis [109] and the retina [38, 128]. Apart from clinical use, OCT also has applications in animal studies with the advantage of repetitive biopsies at one animal at different time points instead of lethal biopsies at different animals for each time point. The particular objective of this study is the morphometric analysis of the skin in the vicinity of a percutaneous implant situated in the lateral abdominal region of a hairless mouse to draw conclusions on its biocompatibility (see Fig. A.1a).

A.1.1 Problem Setup

OCT measures the backscattering profile of a light beam penetrating the sample in axial direction. We use in our setting a spectral-domain OCT (SD-OCT) which enables a shorter acquisition time since it acquires the backscattering profile in spectral domain rather than time domain. For 3D volume acquisition, single axial scan (A-scan) acquisition is combined with a lateral scanning mechanism. 2D scans (B-scans) are composed by a series of A-scans along the x-axis (fast scanning axis). 3D volume scans in turn consist of a series of B-scans along the y-axis (slow scanning axis). In our setting we are confronted with severe axial motion shift due to heart beat or breathing during in vivo SD-OCT (spectral-domain OCT) volume acquisition of mouse skin tissue around a percutaneous implant (see Fig. A.3). Along the fast scanning axis, motion artifacts are illustrated by averaging three consecutive B-scans, resulting in noticeable image blur. In slow scanning direction, motion ar-

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.



(a) (b)

Figure A.2: Image distortion effect in OCT scans. (a) Single OCT B-scan (cropped at half) showing the distorted baseline, and (b) corresponding undistorted result using our method.

tifacts manifest itself in dithering in axial direction as is illustrated in the bottom of Fig. A.3.

As the optical properties of the tissue introduce distortions into the OCT images [121], segmentation based image undistortion is an important step towards fully automatic image analysis tasks.

A.1.2 Motion Compensation

Yun et al. [130] have investigated motion artifacts of SD-OCT occurring during a single A-scan capturing such as signal fading, spatial distortion and blurring. These artifacts can be reduced by increasing the A-scan acquisition rate. However, image shifts in axial direction of several pixels occurring during acquisition of several thousands of A-scans (e.g. for volume acquisition) are still an issue. Later works [96, 75] focus on compensating such image shifts in full volume scans using reference measures. While Ricco et al. [96] compensate transverse motion in retinal volume scans using scanning laser ophthalmoscopy (SLO) images as a reference measure, Lee et al. [75] correct motion shift in dynamic SD-OCT imaging, periodically capturing the same region over several seconds, using one of such captures as reference. Recent work in [68] correct motion artifacts by estimating a displacement vector for each A-scan using orthogonal OCT scan patterns. The method works without having a reference measure. Inference of the displacement field is done by minimizing an objective function using a gradient-descent method combined with a multi-resolution approach. Our work extends this approach by transferring the objective function to CRF notation and adding additional priors, allowing better tissue structure preservation and fast global optimization.

Contributions: We propose a probabilistic method for estimation and compensation of axial motion shift in *in vivo* SD-OCT without requiring a reference measure. The key challenge is to distinguish between motion shift and the natural spatial structure of the subject tissue. We tackle this problem by combining two different lateral scanning schemes for volume acquisition: The motion shift of multiple taken

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

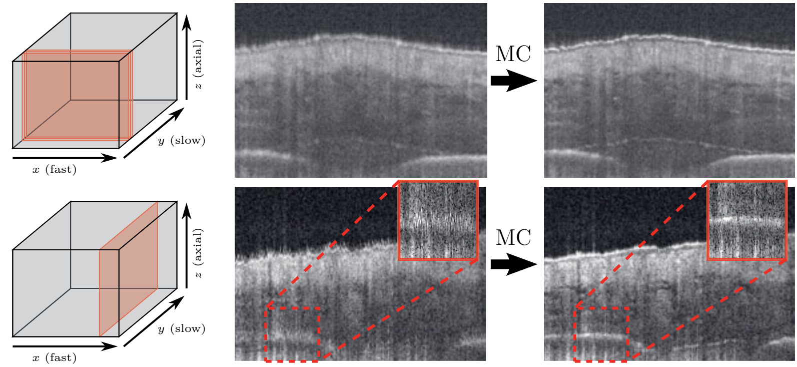


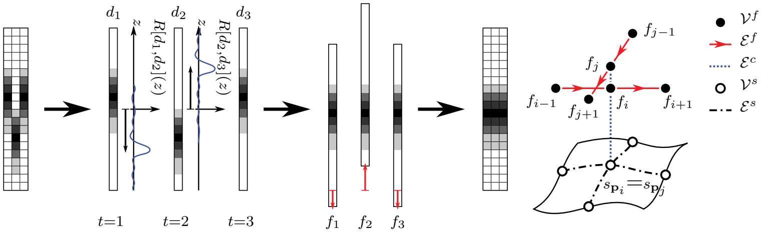
Figure A.3: In vivo SD-OCT scan (cropped) and its motion compensated (MC) result. Top: Slice along the fast scanning axis. Three slices are averaged for visualization of the motion distortion. Bottom: One reslice along the slow scanning axis.

A-scans at the same lateral position (but at various time points) differ whereas the tissue structure remains unchanged. The motion compensation problem is formulated as an energy minimization problem using a conditional random field (CRF) notation, allowing both estimation of the motion field and the tissue structure. For inference, the CRF is simplified to a Gaussian Markov random field (GMRF) by approximating crosscorrelation terms with a Gaussian pdf. Finally, our method is applied on in vivo SD-OCT scans of skin tissue with a percutaneous implant (see Fig. A.3, dashed red rectangles indicate the subcutaneous implant base).

Motion Field Model

For estimation and compensation of in vivo subject movement, the following assumptions are made: A sequence of A-scans $\{d_t\}$ is captured at discrete time points t and lateral position $\mathbf{p}_t = (x_t, y_t)$. For image acquisition, it is assumed that the scanned subject is somehow fixed (e.g. no freehand capturing involving transverse motion drift). Nevertheless, subject movements can not be suppressed completely, e.g., slight up-down movements caused by breathing or heart beating can still occur, but are limited in amplitude. Thus, for each A-scan d_t, we have a corresponding axial *motion shift* f_t. Since axial shift is not solely determined by f_t due to spatial tissue structure changes, the true axial shift is defined by $f_t + s_{\mathbf{p}_t}$, where $s_{\mathbf{p}_t}$ is the tissue surface height at lateral position $\mathbf{p}_t$. In the following, we derive a CRF model $E(f{,}s \mid d) = E(d \mid f{,}s) + E(f{,}s)$ given an observation model $E(d \mid f{,}s)$ and regu-

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.



(a) (b)

Figure A.4: Proposed model: (a) Motion correction workflow: Motion field $\{f_i\}_i$ (red arrows) is estimated by maximizing the crosscorrelation $R[d_i,d_j](z = f_i - f_j)$ (blue curves) of adjacent image rows d_i, d_j, (b) Graphical model structure (red arrows indicate temporal scanning direction).

larizer $E(f,s)$ with $f = \{f_t\}$ as described below. $E(f,s \mid d)$ is defined over a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ given the vertex set $\mathcal{V}$ containing model instances and the edge set $\mathcal{E}$ representing dependencies between instances.

Observation model: The observation model is based on the assumption of structural similarity of (i) spatial neighbored A-scans d_i and d_j with $(i,j) \in \mathcal{E}^R$, where the structure of $\mathcal{E}^R$ depends on the scanning scemes used (see Fig. A.5c in Sect. A.1.2) and (ii) A-scans taken at the same spatial position but at different time points d_i and d_j with $(i,j) \in \mathcal{E}^c = \{(i,j) \mid \mathbf{p}_i = \mathbf{p}_j\}$. As a similarity measure of adjacent A-scans, we use the crosscorrelation $R[\tilde{d}_i, \tilde{d}_j](z) = \int \tilde{d}_i(\tau) \cdot \tilde{d}_j(z+\tau)\,\mathrm{d}\tau$ of two adjacent volume gradient columns $\tilde{d}_i, \tilde{d}_j$, with $z = f_i - f_j$ denoting the relative motion shift (see Fig. A.4a). Actually, the axial shift does not only depend on the motion shift itself, but also on the spatial change of the tissue surface structure. Therefore, we introduce new model variables $\{s_{\mathbf{p}_i}\}_{\mathbf{p}_i \in \mathcal{V}^s}$ denoting the subject surface change. Then, the relative axial shift is now determined by $f_i + s_{\mathbf{p}_i}$, rather than only by f_i, i.e. z becomes $(f_i + s_{\mathbf{p}_i}) - (f_j + s_{\mathbf{p}_j})$. If s is not known a priori, the problem is ill-posed because of ambiguities in the sum of relative motion and surface change. We solve this ambiguity by correlating A-scans taken at the same sample position at different time points. For such A-scans, it is $z = f_i + s_{\mathbf{p}_i} - f_j - s_{\mathbf{p}_j} = f_i - f_j$, because $s_{\mathbf{p}_i} = s_{\mathbf{p}_j}$ $\forall (i,j) \in \mathcal{E}^c$. Thus, we have

$$E(d \mid f, s) = \gamma \sum_{(i,j) \in \mathcal{E}^R} R_{ij}(f_i + s_{\mathbf{p}_i} - f_j - s_{\mathbf{p}_j}) + \sum_{(i,j) \in \mathcal{E}^c} R_{ij}(f_i - f_j) \qquad \text{(A.1)}$$

where $R_{ij}(\,\cdot\,) := -\log R[\tilde{d}_i, \tilde{d}_j](\,\cdot\,)$ and γ is a weighting factor.

Motion field prior: For regularizing the motion estimation problem, additional assumptions are encoded in the prior energy term. Due to mass inertia of the

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

subject, the motion field has to be smooth in time direction. In our model, we use first order smoothness. Additionally, we assume a Gaussian motion shift prior with zero mean, i.e. $f_t \sim \mathcal{N}(0, \sigma_f^2)$. The tissue surface s is regularized analogously. Thus, the prior is:

$$E(f,s) = \theta_1 \sum_{i \in \mathcal{V}^f} \frac{f_i^2}{2} + \theta_2 \sum_{(i,j) \in \mathcal{E}^f} (f_i - f_j)^2 + \theta_3 \sum_{i \in \mathcal{V}^s} \frac{s_{\mathbf{p}_i}^2}{2} + \theta_4 \sum_{(i,j) \in \mathcal{E}^s} (s_{\mathbf{p}_i} - s_{\mathbf{p}_j})^2 \quad \text{(A.2)}$$

with $\mathcal{E}^f = \{(t, t+1) \mid t \in [0, 1, \ldots, T]\}$ and $\theta_1 = 1/\sigma_f^2$, $\theta_2 = \lambda_f$, $\theta_3 = 1/\sigma_s^2$, $\theta_4 = \lambda_s$, $\theta_5 = \gamma$ are the model parameters. The composed graph structure $\mathcal{G} = (\mathcal{V}^f \cup \mathcal{V}^s, \mathcal{E}^f \cup \mathcal{E}^R \cup \mathcal{E}^s \cup \mathcal{E}^c)$ is depicted in Fig. A.4b.

Inference

To efficiently find a configuration $\{f^*, s^*\}$ minimizing $E(f, s \mid \theta)$, we have decided to simplify $E(f, s \mid \theta)$. The only terms in $E(f, s \mid \theta)$ which makes efficient inference difficult are the crosscorrelations $R[d_i, d_j](\,\cdot\,)$. Assuming that the A-scan intensities follow an edge-step model and is augmented with additive white Gaussian noise, the crosscorrelation of the first derivative of the A-scan intensities has a Gaussian shape with additive white Gaussian noise. For model parameter estimation, nonlinear least-squares Gaussian fitting is applied. Thus we obtain $R[d_i, d_j](z) \approx \hat{R}[d_i, d_j](z) = \mathcal{N}(\mu_{ij}, \sigma_{ij}^2, z)$, where μ_{ij} and σ_{ij}^2 are the mean and variance of the estimated Gaussian distribution $\mathcal{N}(\mu, \sigma^2, \,\cdot\,)$.

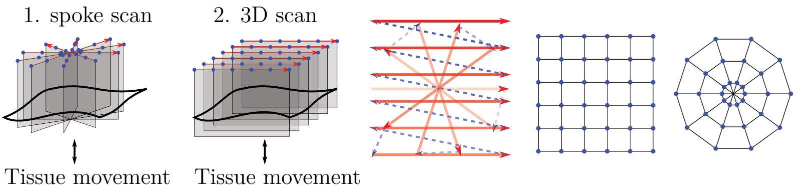
Using this approximation, the CRF energy function $E(f, s \mid d)$ simplifies to a Gaussian Markov random field (GMRF), i.e. $E(f, s \mid d)$ is a quadratic function in $\{f, s\}$ and can be rewritten as $E(x \mid \theta) = \frac{1}{2} x^{\mathsf{T}} A_\theta x + x^{\mathsf{T}} b_\theta + c_\theta$, where $x = \{f, s\}$ and A_θ is sparse due to the Markov property. Its minimizer is $x^* = -A_\theta^{-1} b_\theta$. This can be efficiently solved by (sparse) Cholesky decomposition of A_θ [97].

Estimation of the optimal parameter vector θ^* is done by minimizing the mean-square-error (MSE) of f with $\theta^* = \arg\min_\theta \|f_\theta^* - f_{\text{correct}}\|_2^2$, where $x_\theta^* = \arg\min_x E(x \mid d, \theta)$ with $x_\theta^* = \{f_\theta^*, s_\theta^*\}$ and f_{correct} is the ground truth motion field. In practice, it is sufficient to set σ_s, λ_s and γ fixed (e.g. $\sigma_s = 500$, $\lambda_s = 0.01$ and $\gamma = 1$) and only optimize over σ_f and λ_f, since the former parameters don't affect the estimation results much. Finally, grid search is performed for estimation of $\sigma_f \in \{10, 100, 1000, 10000\}$ and $\lambda_f \in \{0.0001, 0.001, 0.01, 0.1\}$.

Experiments and Discussion

In this section, we compare three different settings of our proposed method. The first setting uses $\gamma = 1$, $\sigma_s = 500$ and $\lambda_s = 0.01$. In the second setting, the tissue surface is ommited, enforcing $s \equiv 0$, i.e. $\sigma_s \to 0$ and $\lambda_s \to \infty$. The third setting additionally omits the spatial crosscorrelation ($\mathcal{E}^R$) term, i.e. $\gamma = 0$, leading to a configuration most similar to the approach of Kraus et al. [68].

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.



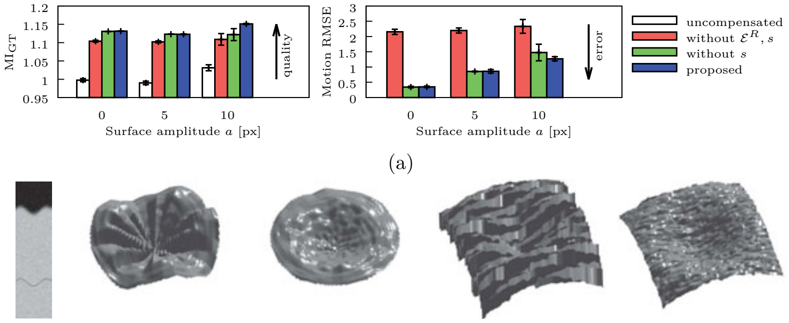
(a) (b) (c)

Figure A.5: Scanning schemes: (a) spoke and dense 3D pattern, (b) lateral positions of B-scans (red lines) over time. Dashed blue lines: connection of consecutive B-scans. Color shading and arrows depict temporal direction, (c) spatial structure of $\mathcal{E}^R$ for spoke and 3D pattern. Blue points: A-scans d_i, black lines: neighborhood relations $(i,j) \in \mathcal{E}^R$.

We present two different experiments involving synthetic data and real OCT acquisitions. The first experiments are done on synthetic datasets, where ground truth surface and motion fields are available. In a second part, real OCT scans of both post mortem (with artificial motion field) and in vivo (without prior known motion field) are evaluated. For synthetic data, as well as real OCT measures, the subjects are scanned consecutively with two scanning schemes for ensuring that enough surface points are scanned twice. The first scheme is a spoke pattern scanning scheme with N_{spoke} B-scans, each B-scan consists of N_{A} A-scans as shown in Fig. A.5a. The second scanning scheme is a dense 3D (cuboidal) scanning scheme with N_{3D} B-scans. Figure A.5b shows a schematic of the lateral scanning positions over time of a complete subject scan. Figure A.5c shows the spatial structure of neighboring, crosscorrelated A-scans (encoded in $\mathcal{E}^R$).

Synthetic data is generated with $N_{\text{spoke}} = 16$, $N_{\text{3D}} = 100$, $N_{\text{A}} = 100$ and axial resolution of $Z = 600\,\text{px}$. The tissue is modeled as a uniformly scattered medium with the tissue-air interface modeled by a step edge function convolved with a Gaussian kernel with $\sigma_{\text{step}} = 5$. The image intensities (with range $[0,1]$) are corrupted with additive Gaussian noise with $\sigma^2_{\text{noise}} = 0.07$. Artificial motion artifacts were generated by adding two sine waves of random amplitude and phase to simulate periodic movement. Low-frequency random shift of up to ±20 px is added for simulation of non-periodic movement. We evaluated our motion compensation algorithm on data with sinusoidal tissue surface of amplitude a as shown in Fig. A.6b. Performance evaluation is done using mutual information (MI) inspired by [68], i.e. measuring the similarity between spoke scan volume and 3D scan volume (resliced to capture the same regions as the spoke scan), denoted with $\text{MI}_{\text{sp-3D}}$. Since ground truth volumes for synthetic data is available, we can also compute the MI of the ground truth volume scans to its motion compensated volume, denoted with

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.



(a)

(b) (c) (d)

Figure A.6: Synthetic data results: (a) Evaluation of the mutual information towards the ground truth (MI_{GT}, left) and the RMSE of the motion fields towards ground truth for different surface amplitudes (right) and (b)–(d) example surface segmentation of a synthetic dataset with non-planar surface. (b) Example B-scan slice, (c) Spoke scans and (d) dense 3D volume scans. Left: with motion artifacts, right: motion compensated.

MI_{GT}. Figure A.6a shows MI_{GT} and motion RMSE results of 30 randomly generated datasets with varying tissue surface amplitudes (10 datasets for $a = 0$, 5, and 10 respectively) with errorbars indicating the standard deviation. The results show best performance on the first configuration, showing most increase of MI and least motion RMSE. The configuration ommiting $\mathcal{E}^R$ and s performs worst on every dataset. $MI_{\text{sp-3D}}$ gives nearly similar results for every configuration, since this measure only captures intra-volume similarity enforced by the $\mathcal{E}^c$ term and cannot capture the tissue structure preservation, as noticed in [68]. Figure A.6c–(d) shows a comparison of extracted tissue surface renderings of uncompensated to compensated volumes.

Real OCT scans: Our real world application uses in vivo and post mortem SD-OCT scans of the percutaneous implant of an anesthetized (and fixed) mouse from [84]. The setting has following parameters: $N_{\text{spoke}} = 72$, $N_{3D} = 800$, $N_A = 800$ and an axial resolution of $Z = 600\,\text{px}$. Acquisition time was approx. 0.1 s per B-scan. For enhancement of computation time and memory usage, a downsampling along the fast scanning axis by a factor of 8 is applied and the motion field is upsampled afterwards for providing motion compensation in full resolution. In Fig. A.7, the evaluation results of one post mortem dataset (p. m.) corrupted with artificial motion (thus known ground truth) and several in vivo datasets (without known ground truth) are shown. For both post mortem and in vivo scans a increase of MI_{GT} and $MI_{\text{sp-3D}}$, respectively, is observed, showing a significant reduction of

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

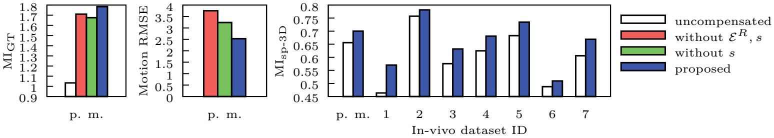


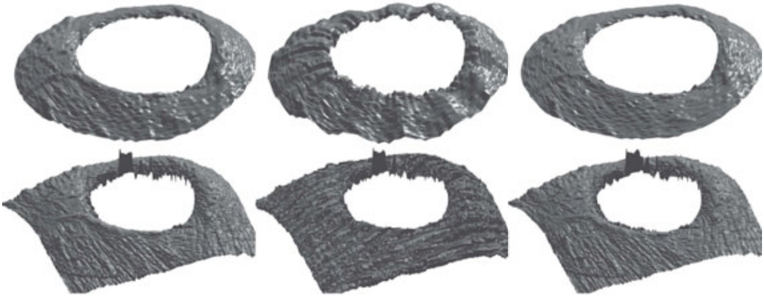
Figure A.7: Evaluation on post mortem data (with known ground truth) using $\mathrm{MI}_{\mathrm{GT}}$ and motion RMSE and in vivo data (without known ground truth) using $\mathrm{MI}_{\text{sp-3D}}$.

motion artifacts. This finding can also be observed in the surface segmentation visualization of the post mortem dataset (see Fig. A.8a) and a typical in vivo data example as shown in Fig. A.8b and Fig. A.3.

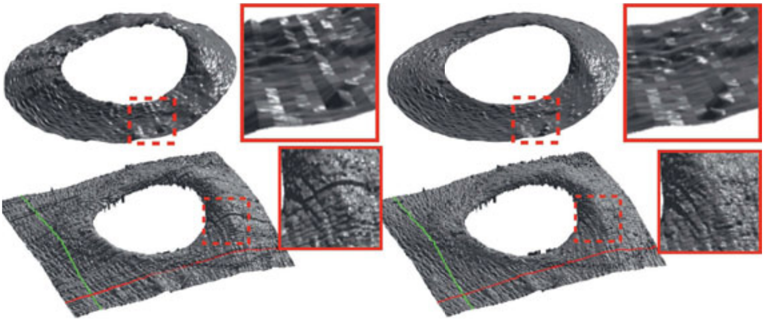
Conclusion

In this work, we propose a novel probabilistic approach for motion compensation of in vivo SD-OCT volume scans. The motion estimation problem is reformulated as a CRF energy function and approximated by a GMRF for efficient inference. Our method reliably separates axial motion from tissue structure change by combining two scanning schemes. We use multiple A-scans taken at the same lateral position but different time points as anchor points to estimate the tissue morphology. The method is verified on synthetic data as well as in vivo SD-OCT volume scans. Motion artifacts are significantly reduced while the geometry of the tissue is preserved.

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.



(a)



(b)

Figure A.8: Motion compensation results: (a) post mortem dataset with ground truth (left), artificial motion (middle), and motion compensated (right) and (b) in vivo data with motion artifacts (left) and motion compensated (right). Red and green lines indicating the position of slices and reslices respectively shown in Fig. A.3. Top row: spoke pattern scan, bottom row: dense 3D scan respectively.

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

A.1.3 OCT Image Undistortion

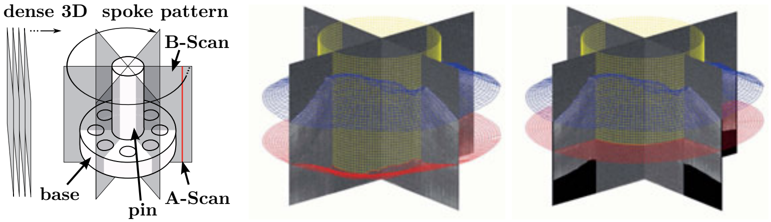
As the optical properties of the tissue introduce distortions into the OCT images [121], segmentation based image undistortion is an important step towards fully automatic image analysis tasks. In recent works, several methods like graph based global optimization, active contours and random fields are proposed for layer segmentation. In practice, graph based approaches, as used in [40] for fully automatic 3D retinal multilayer segmentation, lead to huge graph sets, limiting the number of voxels. Active contour models (e.g. snakes) [59, 128] provide robust results, however require manual initialization. In [58], a fully automatic 2D feature segmentation is presented using conditional random fields and efficient optimization algorithms for inference. Segmentation of a single 2D OCT scan (B-scan) can be susceptible to local shading effects and image perturbations and extending the scanning scheme to the third dimension can significantly improve the segmentation quality [33, 40, 49]. Thus our algorithm is based on 3D segmentation.

This paper proposes an approach for fully automatic segmentation of 3D Fourier-domain OCT and refractive undistortion. The determination of the refractive index is facilitated by the geometry of the implant which consists of a percutaneous pin (3 mm diameter and 5 mm length) anchored beneath the dermis by a flat disc shaped base which is visible in OCT (see Fig. A.1a–(b)).

Two main technical **contributions** are proposed. First, estimation of the skin surface in the 3D space from several OCT B-scans is done using a Markov random field (MRF) approach with an efficient combination of global and local optimization algorithms. A spoke pattern scanning scheme is used for 3D data acquisition and is further compared with a dense 3D scanning scheme (see Fig. A.9a). Our second contribution addresses the segmentation of the implant base. The distorted implant base is segmented using a refractive distortion model and the previously segmented skin surface for parameter estimation in order to match the distorted implant base best to the a priori known shape of the undistorted base contour. The parameters of the implant base are estimated with a fast generalized 3D Hough transform approach, optimizing the refractive index, as well as the 3D position and orientation of the implant base. The segmented model is finally used for refractive image undistortion (see Fig. A.2a–(b)).

In Section 2, the 3D segmentation of the pin position, the skin surface and the implant base is described (Fig. A.9b), which is used for refractive image undistortion (Fig. A.9c). In Section 3, the used undistortion model is verified and a comparison of the spoke pattern and dense 3D scanning scheme is shown, followed by a quantitative analysis of several mouse datasets segmented and undistorted using our method. Finally, in Section 4, a short conclusion is given.

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.



(a) (b) (c)

Figure A.9: (a) Schematic of the percutaneous implant (pin and base) and OCT scanning schemes (dense 3D and spoke pattern), (b) two orthogonal OCT B-scans with segmentation (mesh overlay) of pin position (yellow), skin surface (blue) and deformed base (red), (c) undistorted B-scans with mesh overlay.

Methods

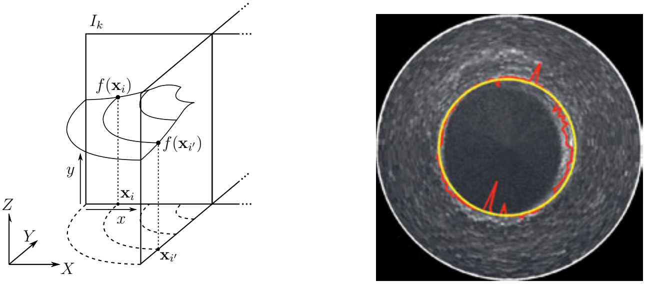
The OCT data is acquired in a sequence of B-scans $(I^k)_{k=1,\ldots,K}$ (see Fig. A.9a) with image width W and height H. To reduce noise and small scanning artifacts, while preserving edges, we apply a median filter to each B-Scan as a first preprocessing step. In a second step we apply a pixel intensity normalization to each image, leading to a zero-mean intensity distribution with unit variance, in order to retrieve uniform edge responses from the in the final preprocessing step applied edge filter. We use a Sobel filter in combination with a presmoothing Gaussian kernel with $\sigma_{\mathrm{gauss}} = 1.5$ to get first order derivative images I^k_{x} and I^k_{y} in x- and y-direction of I^k.

For 3D segmentation, the two-dimensional B-scans are embedded in a global 3D coordinate system. A mapping of a 3D point position $\mathbf{P} = (X, Y, Z)$ to image coordinates $\mathbf{p} = (x, y)$, i.e. $(X, Y, Z) \mapsto (k, x, y)$ as shown in Fig. A.10a is done, resulting in a sparse volume representation $V(X, Y, Z) = I^k(x, y)$ for the image intensities. The volumes of the image derivatives V_{x}, and V_{y} are analogously defined, using I^k_{x} and I^k_{y} instead of I^k.

The proposed implant segmentation method is divided into three consecutive steps (see Fig. A.9b–A.9c): The pin segmentation (yellow cylinder), the skin surface segmentation (blue mesh), and the base segmentation (red mesh). The segmentation steps are described in the following subsections.

Pin Segmentation In the image area where the implant pin is located, there is no contour information, thus this area can be ignored for following segmentation steps. The pin is of cylindrical form and the diameter d is known a priori. The diameter is allowed to have a variance v^k of $\pm 10\,\mathrm{pel}$. For each image I^k, the x-position of the left and right pin boundary x^k_1 and $x^k_2 = x^k_1 + d + v^k$ is computed using a generalized Hough transform [7] approach over $I^k_{\mathrm{acc}}(x) = \sum_y |I^k_{\mathrm{x}}(x, y)|$ with

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.



(a) (b)

Figure A.10: (a) Modeling of skin surface f over grid points $\mathbf{x}_i, \mathbf{x}_{i'}$ with respect to global coordinate system (X, Y, Z) using spoke pattern scanning scheme, (b) top view of accumulated image intensities I_{acc} and segmented pin boundary for each B-scan individually (red), and with fitted cylinder (yellow).

$\max_{x_1^k, v^k} \left[I_{\text{acc}}^k(x_1^k) + I_{\text{acc}}^k(x_1^k + d + v^k) \right]$. The pin estimation in a single B-scan is susceptible to noise and vanishing of boundary contours (see Fig. A.10b, red line). Therefore, the boundaries are smoothed by assuming a cylindrical form and radius r of the pin: At first, the center position $\mathbf{C} = (C_X, C_Y)$ of the pin in the XY-plane is calculated as the arithmetic mean of all boundary positions $\mathbf{B}_1^k$ and $\mathbf{B}_2^k$ corresponding to x_1^k and x_2^k, i.e. $\mathbf{C} = \frac{1}{2K} \sum_k (\mathbf{B}_1^k + \mathbf{B}_2^k)$. Then, the radius r is computed as the median of all radians $r = \text{median}(|\mathbf{CB}_1^1|, \ldots, |\mathbf{CB}_1^K|)$. The yellow line in Fig. A.10b shows a typical result of the pin segmentation step.

Skin Modeling and Segmentation A correct estimation of the skin surface is crucial for a correct modeling of the implant base. The skin surface, denoted as $f(X, Y) = Z$, is assumed to be smooth and to behave like a membrane, thus having no discontinuities, except the pinhole. Several approaches can be used to model the surface: Markov Random Fields (MRF), Conditional Random Fields or Discriminative Random Fields [76]. Due to the smoothness property, we decided to use an adapted MRF for segmentation of the skin surface. Following the notations in [76], the posterior probability of the skin surface f given the volumetric pixel intensities V can be written as $P(f \mid V) \propto P(f, V) = p(V \mid f)P(f)$ using the Maximum A-Posteriori framework (MAP), where $P(f)$ is the smoothness prior and $p(V \mid f)$ denotes the likelihood function of f for V fixed.

To represent probabilistic relationships of the MRF, a common graphical notation is used. The neighborhood relationships of a MRF can be described given a graph

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

$G = (V,E)$ with a set of nodes V representing the instances of the random field and a set of edges E representing the conditional dependencies between the instances. The prior probability is then modeled as follows:

$$P(f) = \exp\left[-\textstyle\sum_{i \in V} U(f_i)\right] \quad \text{with} \qquad U(f_i) = \textstyle\sum_{\{i,i'\} \in E} (f_i - f_{i'})^2 / 2\sigma_s^2 \ , \tag{A.3}$$

where σ_s is a constant weighting factor. Figure A.10a shows the relation of the skin surface $f_i = f(\mathbf{x}_i)$ in the observation point $\mathbf{x}_i$ to its neighboring point $\mathbf{x'}_i$. A 4-neighborhood system is used.

The likelihood of the true skin surface at position $\mathbf{x}_i = (x_i, y_i)$ going through the volumetric point (X_i, Y_i, Z_i), with $Z_i = f(\mathbf{x}_i)$ is given as

$$p(V|f_i) \propto V_{\mathrm{y}}(X_i, Y_i, f(\mathbf{x}_i)) + c \ , \tag{A.4}$$

with a shifting constant c, forcing strict positive probabilities. Determining the optimal solution of the given MRF problem is transferred to finding the global maximum of $P(f \mid V)$. The given MRF model consists of $K \cdot W$ edges and $K \cdot W \cdot H$ observation points, thus searching for the global maximum of the joint probability turns out to be a complex task. In order to solve this task in a reasonable amount of time and memory usage, we decided to use an iterative local optimization algorithm. The Iterated Conditional Modes (ICM) approach with the *coding method* of Besag [13] is used because of its ability for fast convergence. The ICM algorithm searches for a local maximum of the joint probability $P(f \mid V)$ by iteratively maximizing each local probability $P(f_i \mid V)$ independently:

$$f_i^{n+1} = \arg\min_z\left[-\log p(V \mid f_i^n = z) + \textstyle\sum_{\{i,i'\} \in E} (z - f_{i'}^n)^2 / 2\sigma_s^2\right] \ , \tag{A.5}$$

where f^0 is an initial guess of the surface. ICM is a local minimization method and the estimation result highly depends on the initial surface guess f^0. Therefore, the initial guess is retrieved by independently estimating an optimal path for each B-scan using a Markov model, i.e. the same model as for the MRF, but with a 2- instead of a 4-neighborhood system. Finally, the Viterbi algorithm is used for global optimization [113]. Additionally, an annealing procedure inspired by annealing labeling ICM of [76] is used, i.e. allowing the membrane for $n = 0$ to be more relaxed by setting σ_s^0 to a higher value σ_{start} and decreasing it for each L's iteration by $\sigma_s^{n+L} = \max\{\sigma_{\mathrm{end}},\, \sigma_s^n \cdot \sigma_{\mathrm{decr}}\}$.

Baseline Modeling and Detection with a Hough Transformation To achieve an appropriate segmentation of the implant base in presence of scanning artifacts, noise, and local vanishing edge structures, a robust and model based segmentation approach was developed using a generalized Hough transform [7]. The applied refractive image undistortion model uses the fact that the axial position of reflections captured with the OCT system matches to the optical path length z_{opt} of light

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

passing through the observed tissue, rather than the geometric path length z_{geo}. Inspired by [109], the relation between z_{opt} and z_{geo} can be approximately formulated as $z_{\text{opt}} = n z_{\text{geo}}$, as shown in Fig. A.2a–(b), assuming a homogeneous layer with refractive index n. The model used for conversion of optical path length to geometric path length of each axial scan (A-scan, see Fig. A.9a) is $z = g_u + (g_l - g_u)/n$, where g_u is the known upper position (the skin surface, estimated in Section A.1.3), g_l the lower position (base layer), and z denotes the geometric position of the base layer (see Fig. A.2a). Given g_u and the constraint of maximal edge intensity support of I_{y} over g_l, the maximization term for the generalized Hough transform is stated as follows:

$$\max_{Z,n,\theta_X,\theta_Y} \textstyle\sum_{x,k} I_{\text{y},k}\left(x, g_u(x,k) - \left[g_u(x,k) - Z_{\theta_X,\theta_Y}(x,k)\right] \cdot n\right) \quad . \tag{A.6}$$

The implant base is modeled as a plane with geometric position Z. Since the pin is not located perfectly horizontal, a rotation of the plane in X- and Y-direction is applied to Z, denoted by Z_{θ_X,θ_Y}. As the parameter space is of dimension 4, small discretization step sizes lead to high computation time, i.e. doubling the precision increases the computation time by a factor of 2^4. Since the minimization problem has only one global optimum, which can be distinguished very well from small local extrema, a resolution pyramid approach [52] is applied.

Experiments

In this section, a ground truth experiment for verification of the proposed undistortion model, as well as a comparison of the used spoke pattern and dense 3D scanning scheme is performed. We further show results of a quantitative study on several mouse datasets[1]. The B-scans have a dimension of 800 px × 600 px with a lateral distance of the A-Scans of 7.5 µm/px and an axial scale of 4.7 µm/px. Following parameters work best for our datasets: An 11×11 median filter, $\sigma_{\text{start}} = 70$, $\sigma_{\text{end}} = 10$, $\sigma_{\text{dec}} = 0.9$, and $K = 5$.

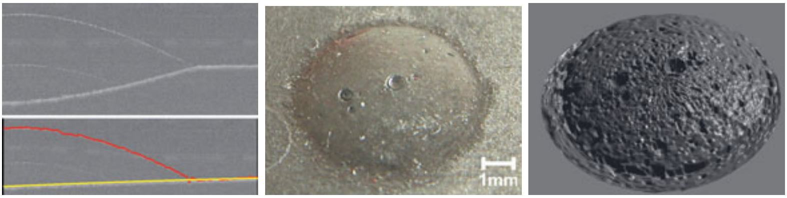
Model Verification A plane plastic slide is prepared with two glue drops of slightly different size (Vitralit® 4731) with known refractive index of $n = 1.474$. An example image and its segmentation results are shown in Fig. A.11. The estimated refractive index of the two glue drops are $n_{\text{est}} = 1.494$ and $n_{\text{est}} = 1.507$ respectively. It is assumed that the ground carrier plate is not perfectly planar as expected by the estimation model, causing the deviation from groundtruth.

Scanning Scheme The spoke pattern scanning scheme is compared with the dense 3D scanning scheme using real mouse recordings. To this end, we use the skin

---

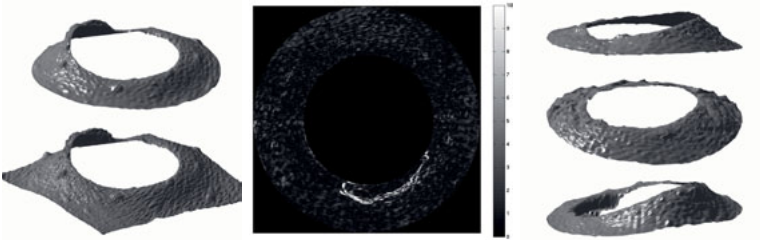
[1] All animal experimental procedures have been approved by the local governmental animal care committee (Approval No. 33-42502-04-08/1498).

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.



(a) (b) (c)

Figure A.11: Glue drop (Vitralit® 4731) with $n = 1.474$, (a) comparison of original (top) with segmented and undistorted (bottom) OCT B-scan (images cropped), (b) closeup photo of a glue drop, (c) corresponding rendered surface reconstruction.



(a) (b) (c)

Figure A.12: (a) Comparison of spoke pattern scanning scheme (top) and dense 3D scanning scheme (bottom), (b) comparison heightmap of absolute surface differences in pel, (c) example skin segmentation results of mouse datasets using spoke pattern.

surface reconstructions of scans captured using the spoke pattern with 72 B-scans and the dense 3D scanning scheme, respectively. Figure A.12a shows reconstructions for a typical skin surface (acquired post mortem). The segmented surface using the spoke pattern is projected onto the dense 3D grid. Small surface differences show the ability of capturing important surface features using the spoke pattern (see Fig. A.12b). The root mean square deviation (RMSD) is calculated as 1.546 pel. Furthermore, several other surfaces were reconstructed (see Fig. A.12c). The results show, that using the spoke pattern still leads to good results and preserves many skin details. Moreover, scanning is faster and approximately 91 % of computation time and disk space is saved due to the decreased number of B-scans. With 937.5 KiB per image, 666.5 MiB are saved. Skin surfaces of spoke pattern scans are reconstructed using an unoptimized MATLAB implementation in 5.78 min, compared to 68.14 min using dense 3D scans.

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

Quantitative Study We further carried out a quantitative analysis on a set of 60 OCT mouse scans of 23 mice at various points of time (including post mortem scans) using the spoke pattern scanning scheme. Manual segmentations of the skin and base were done from experts for 8 OCT images (each 9th slice) per scan, having 3 individual manual segmentations per slice. The experts were instructed to trace only visible parts of the contours. As a metric, we use the RMSD of a surface S_1 towards the mean of a set of surfaces $S_2, \ldots, S_m$. For each B-scan, the RMSD of the automatic skin segmentation towards the mean of all manual skin segmentations is calculated. The average and standard deviation (in pel) over all B-scans is 3.98 ± 3.29. For comparison, the RMSD of each manual skin segmentation towards the mean of all other manual skin segmentations is calculated with 3.62 ± 1.03. For automatic vs. manual base segmentations, we get: 12.90 ± 18.27 and 3.55 ± 4.51. After outlier removal (B-scans with RMSD $>=$ 10 pel), the RMSD of our fully automated approach is 3.63 ± 1.33 (2.3 % outlier) and 5.27 ± 2.40 (39.0 % outlier), which is close to the RMSD of the manual segmentation performed by experts with 3.62 ± 1.00 (0.1 % outlier) and 3.21 ± 1.37 (1.4 % outlier). Outliers are mostly due to motion artifacts in scans captured in vivo. Future work will concentrate on reducing the outliers.

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte
Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

Bibliography

[1] Ijaz Akhter and Michael J. Black. Pose-conditioned joint angle limits for 3D human pose reconstruction. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1446–1455, 2015.

[2] Sikandar Amin, Mykhaylo Andriluka, Marcus Rohrbach, and Bernt Schiele. Multi-view pictorial structures for 3d human pose estimation. In *British Machine Vision Conference (BMVC)*, 2013.

[3] Christophe Andrieu, Nando de Freitas, Arnaud Doucet, and Michael I. Jordan. An introduction to mcmc for machine learning. *Machine Learning*, 50(1-2):5–43, 2003.

[4] Mykhaylo Andriluka, Stefan Roth, and Bernt Schiele. Pictorial structures revisited: People detection and articulated pose estimation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009. Best Paper Award Honorable Mention by IGD.

[5] Boris Babenko, Ming-Hsuan Yang, and Serge Belongie. Visual tracking with online multiple instance learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 2011.

[6] Shai Bagon. Matlab wrapper for graph cut, 2006.

[7] Dana H. Ballard. Generalizing the hough transform to detect arbitrary shapes. 13(2):111 – 122, 1981.

[8] Vasileios Belagiannis, Sikandar Amin, Mykhaylo Andriluka, Bernt Schiele, Nassir Navab, and Slobodan Ilic. 3d pictorial structures for multiple human pose estimation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1669–1676, 2014.

[9] Richard E. Bellman. *Dynamic Programming.* Dover Books on Computer Science Series. Dover Publications, 2003.

[10] Jérôme Berclaz, François Fleuret, Engin Türetken, and Pascal Fua. Multiple object tracking using k-shortest paths optimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 2011.

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte
Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

[11] Luca Bertinetto, Jack Valmadre, Stuart Golodetz, Ondrej Miksik, and Philip Torr. Staple: Complementary learners for real-time tracking. *arXiv*, 2015.

[12] Dimitri P. Bertsekas. *Nonlinear Programming*. Athena Scientific, 2nd edition, 1999.

[13] Julian Besag. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society*, 36(2):pp. 192–236, 1974.

[14] Frederic Besse, Carsten Rother, Andrew Fitzgibbon, and Jan Kautz. Pmbp: Patchmatch belief propagation for correspondence field estimation. In *British Machine Vision Conference (BMVC)*, 2012.

[15] Danny Bickson. Gaussian belief propagation: Theory and aplication. *PhD thesis*, abs/0811.2518, 2008.

[16] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3(1):1–122, 2011.

[17] Yuri Boykov and Vladimir Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 26(9):1124–1137, 2004.

[18] Yuri Boykov, Olga Veksler, and Ramin Zabih. Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 23:1222–1239, 2001.

[19] Yuri Y. Boykov and Marie-Pierre Jolly. Interactive graph cuts for optimal boundary & region segmentation of objects in n-d images. In *IEEE International Conference on Computer Vision (ICCV)*, 2001.

[20] Statistic Brain. Youtube company statistics. URL: http://www.statisticbrain.com/youtube-statistics/, Accessed: 01-Sep-2016.

[21] Statistic Brain. Youtube company statistics. URL: http://web.archive.org/web/20120406172003/http://www.statisticbrain.com/youtube-statistics, Accessed: 01-Sep-2016.

[22] Magnus Burenius, Josephine Sullivan, and Stefan Carlsson. 3d pictorial structures for multiple view articulated pose estimation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3618–3625. IEEE Computer Society, 2013.

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

[23] Kai Cordes, Oliver Müller, Bodo Rosenhahn, and Jörn Ostermann. Half-sift: High-accurate localized features for sift. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Workshop*, pages 31–38, 2009.

[24] Kai Cordes, Oliver Müller, Bodo Rosenhahn, and Jörn Ostermann. Bivariate feature localization for sift assuming a gaussian feature shape. In *LNCS International Symposium on Visual Computing (ISVC)*, volume 6453, pages 264–275, 2010.

[25] Kai Cordes, Oliver Müller, Bodo Rosenhahn, and Jörn Ostermann. Feature trajectory retrieval with application to accurate structure and motion recovery. In *LNCS International Symposium on Visual Computing (ISVC)*, volume 6938, pages 156–167, 2011.

[26] Paul Damien, Jon Wakefield, and Stephen Walker. Gibbs sampling for bayesian non-conjugate and hierarchical models by using auxiliary variables. *Journal of the Royal Statistical Society*, 61(2):331–344, 1999.

[27] Martin Danelljan, Andreas Robinson, Fahad Shahbaz Khan, and Michael Felsberg. Beyond correlation filters: Learning continuous convolution operators for visual tracking. In *European Conference on Computer Vision (ECCV)*, 2016.

[28] Jonathan Deutscher, Andrew Blake, and Ian Reid. Articulated body motion capture by annealed particle filtering. In *IEEE Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages 126–133 vol.2, 2000.

[29] Sabine Donner, Oliver Müller, Frank Witte, Ivonne Bartsch, Elmar Willbold, Tammo Ripken, Alexander Heisterkamp, Bodo Rosenhahn, and Alexander Krüger. In situ optical coherence tomography of percutaneous implant-tissue interfaces in a murine model. *Biomedical Engineering/Biomedizinische Technik*, pages 1–9, 2013.

[30] Genquan Duan, Haizhou Ai, Song Cao, and Shihong Lao. Group tracking: Exploring mutual relations for multiple object tracking. In *European Conference on Computer Vision (ECCV)*, pages 129–143, 2012.

[31] Kun Duan, Dhruv Batra, and David Crandall. A multi-layer composite model for human pose estimation. In *British Machine Vision Conference (BMVC)*, 2012.

[32] Kun Duan, Dhruv Batra, and David Crandall. Human pose estimation via multi-layer composite models. *Signal Processing*, 110:15–26, 2015.

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

[33] Justin A. Eichel, Kostadinka K. Bizheva, David A. Clausi, and Paul W. Fieguth. Automated 3d reconstruction and segmentation from optical coherence tomography. In *European Conference on Computer Vision (ECCV)*, ECCV'10, pages 44–57. Springer-Verlag, 2010.

[34] Marcin Eichner, Manuel J. Marin-Jimenez, Andrew Zisserman, and Vittorio Ferrari. 2d articulated human pose estimation and retrieval in (almost) unconstrained still images. *International Journal of Computer Vision (IJCV)*, 99:190–214, 2012.

[35] Ahmed Elhayek, Edilson de Aguiar, Arjun Jain, Jonathan Tompson, Leonid Pishchulin, Mykhaylo Andriluka, Christoph Bregler, Bernt Schiele, and Christian Theobalt. Efficient convnet-based marker-less motion capture in general scenes with a low number of cameras. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.

[36] Pedro F. Felzenszwalb, Ross B. Girshick, David McAllester, and Deva Ramanan. Object detection with discriminatively trained part based models. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 32(9):1627–1645, 2010.

[37] Pedro F. Felzenszwalb and Daniel P. Huttenlocher. Pictorial structures for object recognition. *International Journal on Computer Vision (IJCV)*, 61(1):55–79, 2005.

[38] Delia C. Fernández, Harry M. Salinas, and Carmen A. Puliafito. Automated detection of retinal layer structures on optical coherence tomography images. *Optics Express*, 13(25):10200–10216, Dec 2005.

[39] Vittorio Ferrari, Manuel Marín-jiménez, and Andrew Zisserman. 2d human pose estimation in tv shows. In *In Dagstuhl post-proceedings*, 2009.

[40] Mona K. Garvin, Michael D. Abramoff, Randy Kardon, Stephen R. Russell, Xiaodong Wu, and Milan Sonka. Intraretinal layer segmentation of macular optical coherence tomography images using optimal 3-d graph search. *IEEE Transactions on Medical Imaging*, 27(10):1495–1505, 2008.

[41] Amir Globerson and Tommi S. Jaakkola. Fixing max-product: Convergent message passing algorithms for map lp-relaxations. In *Advances in Neural Information Processing Systems (NIPS)*, pages 553–560. 2008.

[42] Jacques Hadamard. Sur les problèmes aux dérivées partielles et leur signification physique. *Princeton University Bulletin*, 13:49–52, 1902.

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

[43] Jungong Han, Ling Shao, Dong Xu, and Jamie Shotton. Enhanced computer vision with microsoft kinect sensor: A review. *IEEE Transactions on Cybernetics*, 43(5):1318–1334, 2013.

[44] Sam Hare, Amir Saffari, and Philip H. S. Torr. Struck: Structured output tracking with kernels. In Dimitris N. Metaxas, Long Quan, Alberto Sanfeliu, and Luc J. Van Gool, editors, *IEEE International Conference on Computer Vision (ICCV)*, pages 263–270. IEEE, 2011.

[45] Tamir Hazan and Amnon Shashua. Convergent message-passing algorithms for inference over general graphs with convex free energies. *Conference on Uncertainty in Artificial Intelligence (UAI)*, pages 264–273, 2008.

[46] Tamir Hazan and Amnon Shashua. Norm-product belief propagation: Primal-dual message-passing for approximate inference. *IEEE Transactions on Information Theory*, 56(12):6294–6316, 2010.

[47] Tamir Hazan and Raquel Urtasun. A primal-dual message-passing algorithm for approximated large scale structured prediction. In *Advances in Neural Information Processing Systems (NIPS)*, pages 838–846, 2010.

[48] Stefan Holzer, Slobodan Ilic, David Tan, Marc Pollefeys, and Nassir Navab. Efficient learning of linear predictors for template tracking. *International Journal of Computer Vision (IJCV)*, 111(1):12–28, 2014.

[49] Yasuaki Hori, Yoshiaki Yasuno, Shingo Sakai, Masayuki Matsumoto, Tomoko Sugawara, Violeta Dimitrova Madjarova, Masahiro Yamanari, Shuichi Makita, Takeshi Yasui, Tsutomu Araki, Masahide Itoh, and Toyohiko Yatagai. Automatic characterization and segmentation of human skin using three-dimensional optical coherence tomography. *Optics Express*, 14(5):1862–1877, 2006.

[50] Alexander Ihler and David McAllester. Particle belief propagation. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 256–263, 2009.

[51] Alexander T. Ihler. *Inference in sensor networks: graphical models and particle methods.* PhD thesis, 2005.

[52] John Illingworth and Josef Kittler. A survey of the hough transform. *Computer Vision, Graphics, and Image Processing*, 44(1):87–116, 1988.

[53] Michael Isard. Pampas: Real-valued graphical models for computer vision. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, CVPR'03, pages 613–620. IEEE Computer Society, 2003.

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte
Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

[54] Varun Jampani, S. M. Ali Eslami, Daniel Tarlow, Pushmeet Kohli, and John M. Winn. Consensus message passing for layered graphical models. *International Conference on Artificial Intelligence and Statistics (AISTATS)*, abs/1410.7452, 2014.

[55] Vladimir Jojic, Stephen Gould, and Daphne Koller. Accelerated dual decomposition for map inference. In *International Conference on Machine Learning (ICML)*, pages 503–510. Omnipress, 2010.

[56] Zdenek Kalal, Krystian Mikolajczyk, and Jiri Matas. Tracking-learning-detection. *IEEE Transations in Pattern Analysis and Machine Intelligence (PAMI)*, 34(7):1409–1422, 2012.

[57] Jörg H Kappes, Björn Andres, Fred A Hamprecht, Christoph Schnörr, Sebastian Nowozin, Dhruv Batra, Sungwoong Kim, Bernhard X Kausler, Thorben Kröger, Jan Lellmann, et al. A comparative study of modern inference techniques for structured discrete energy minimization problems. *International Journal of Computer Vision (IJCV)*, 115(2):155–184, 2015.

[58] Zahra Karimaghaloo, Mohak Shah, Simon J. Francis, Douglas L. Arnold, D. Louis Collins, and Tal Arbel. Detection of gad-enhancing lesions in multiple sclerosis using conditional random fields. In Tianzi Jiang, Nassir Navab, Josien P. W. Pluim, and Max A. Viergever, editors, *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 41–48. Springer Berlin Heidelberg, 2010.

[59] Michael Kass, Andrew Witkin, and Demetri Terzopoulos. Snakes: Active contour models. *International Journal of Computer Vision (IJCV)*, 1(4):321–331, 1988.

[60] Pushmeet Kohli, Jonathan Rihan, Matthieu Bray, and Philip Torr. Simultaneous segmentation and pose estimation of humans using dynamic graph cuts. *International Journal of Computer Vision (IJCV)*, 79(3):285–298, 2008.

[61] Daphne Koller and Nir Friedman. *Probabilistic graphical models: principles and techniques.* MIT press, 2009.

[62] Daphne Koller, Uri Lerner, and Dragomir Anguelov. A general algorithm for approximate inference and its application to hybrid bayes nets. *Conference on Uncertainty in Artificial Intelligence (UAI)*, abs/1301.6709, 2013.

[63] Vladimir Kolmogorov. Convergent tree-reweighted message passing for energy minimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 28:1568–1583, 2006.

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

[64] Vladimir Kolmogorov and Ramin Zabih. What energy functions can be minimized via graph cuts? *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 26(2):147–159, 2004.

[65] Nikos Komodakis, Nikos Paragios, and Georgios Tziritas. Mrf optimization via dual decomposition: Message-passing revisited. In *IEEE International Conference on Computer Vision (ICCV)*, pages 1–8, 2007.

[66] Nikos Komodakis, Nikos Paragios, and Georgios Tziritas. Mrf energy minimization and beyond via dual decomposition. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 33(3):531–552, 2011.

[67] Rajkumar Kothapa, Jason Pacheco, and Erik B. Sudderth. Max-product particle belief propagation. Technical report, Brown University, 2011.

[68] Martin F. Kraus, Benjamin Potsaid, Markus A. Mayer, Ruediger Bock, Bernhard Baumann, Jonathan J. Liu, Joachim Hornegger, and James G. Fujimoto. Motion correction in optical coherence tomography volumes on a per a-scan basis using orthogonal scan patterns. *Biomedical Optics Express*, 3(6):1182–1199, 2012.

[69] Matej Kristan, Aleš Leonardis, Jiri Matas, Michael Felsberg, and Roman Pflugfelder. Visual object tracking challenge. URL: http://www.votchallenge.net, Accessed: 08-May-2017.

[70] Matej Kristan and Jiri Matas. The visual object tracking vot2016 challenge results, 2016.

[71] Matej Kristan, Jiri Matas, Aleš Leonardis, Michael Felsberg, Luka Čehovin, Gustavo Fernandez, Tomas Vojir, Gustav Häger, Georg Nebehay, Roman Pflugfelder, Abhinav Gupta, Adel Bibi, Alan Lukežič, Alvaro Garcia-Martin, Amir Saffari, Alfredo Petrosino, Andres Solis Montero, Anton Varfolomieiev, Atilla Baskurt, Baojun Zhao, Bernard Ghanem, Brais Martinez, ByeongJu Lee, Bohyung Han, Chaohui Wang, Christophe Garcia, Chunyuan Zhang, Cordelia Schmid, Dacheng Tao, Daijin Kim, Dafei Huang, Danil Prokhorov, Dawei Du, Dit-Yan Yeung, Eraldo Ribeiro, Fahad Shahbaz Khan, Fatih Porikli, Filiz Bunyak, Gao Zhu, Guna Seetharaman, Hilke Kieritz, Hing Tuen Yau, Hongdong Li, Honggang Qi, Horst Bischof, Horst Possegger, Hyemin Lee, Hyeonseob Nam, Ivan Bogun, Jae chan Jeong, Jae il Cho, Jae-Yeong Lee, Jianke Zhu, Jianping Shi, Jiatong Li, Jiaya Jia, Jiayi Feng, Jin Gao, Jin Young Choi, Ji-Wan Kim, Jochen Lang, Jose M. Martinez, Jongwon Choi, Junliang Xing, Kai Xue, Kannappan Palaniappan, Karel Lebeda, Karteek Alahari, Ke Gao, Kimin Yun, Kin Hong Wong, Lei Luo, Liang Ma, Lipeng Ke, Longyin Wen, Luca Bertinetto, Mahdieh Pootschi, Mario Maresca, Martin

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

Danelljan, Mei Wen, Mengdan Zhang, Michael Arens, Michel Valstar, Ming Tang, Ming-Ching Chang, Muhammad Haris Khan, Nana Fan, Naiyan Wang, Ondrej Miksik, Philip H S Torr, Qiang Wang, Rafael Martin-Nieto, Rengarajan Pelapur, Richard Bowden, Robert Laganiere, Salma Moujtahid, Sam Hare, Simon Hadfield, Siwei Lyu, Siyi Li, Song-Chun Zhu, Stefan Becker, Stefan Duffner, Stephen L Hicks, Stuart Golodetz, Sunglok Choi, Tianfu Wu, Thomas Mauthner, Tony Pridmore, Weiming Hu, Wolfgang Hübner, Xiaomeng Wang, Xin Li, Xinchu Shi, Xu Zhao, Xue Mei, Yao Shizeng, Yang Hua, Yang Li, Yang Lu, Yuezun Li, Zhaoyun Chen, Zehua Huang, Zhe Chen, Zhe Zhang, and Zhenyu He. The visual object tracking vot2015 challenge results, 2015.

[72] Frank R. Kschischang, Brendan J. Frey, and Hans-Andrea Loeliger. Factor graphs and the sum-product algorithm. *IEEE Transactions on Information Theory*, 47:498–519, 2001.

[73] L3S. L3s at cebit 2015. URL: https://www.l3s.de/node/800, Accessed: 18-May-2017.

[74] L'ubor Ladický, Philip H. S. Torr, and Andrew Zisserman. Human pose estimation using a joint pixel-wise and part-wise formulation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2013.

[75] Jonghwan Lee, Vivek Srinivasan, Harsha Radhakrishnan, and David A. Boas. Motion correction for phase-resolved dynamic optical coherence tomography imaging of rodent cerebral cortex. *Optics Express*, 19(22):21258–21270, 2011.

[76] Stan Z. Li. *Markov Random Field Modeling in Image Analysis*. Springer Publishing Company, Incorporated, 3rd edition, 2009.

[77] Thibaut Lienart, Yee Whye Teh, and Arnaud Doucet. Expectation particle belief propagation. *International Conference on Neural Information (NIPS)*, abs/1506.05934, 2015.

[78] Zhao Liu, Jianke Zhu, Jiajun Bu, and Chun Chen. A survey of human pose estimation. *Journal of Visual Communitation and Image Representation*, 32(C):10–19, 2015.

[79] Oswaldo Ludwig, David Delgado, Valter Gonçalves, and Urbano Nunes. Trainable classifier-fusion schemes: An application to pedestrian detection. In *IEEE International Conference on Intelligent Transportation Systems (ITSC)*, pages 1–6, 2009.

[80] Talya Meltzer, Amir Globerson, and Yair Weiss. Convergent message passing algorithms - a unifying view. *Conference on Uncertainty in Artificial Intelligence (UAI)*, abs/1205.2625, 2009.

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

[81] Ondrej Miksik, Vibhav Vineet, Patrick Perez, and Philip H. S. Torr. Distributed non-convex admm-inference in large-scale random fields. In *British Machine Vision Conference (BMVC)*, 2014.

[82] Thomas P. Minka. Expectation propagation for approximate bayesian inference. *Conference on Uncertainty in Artificial Intelligence (UAI)*, 2001.

[83] Antonietta Mira. *Ordering, Slicing and Splitting Monte Carlo Markov Chains.* PhD thesis, University of Minnesota, 1999.

[84] Oliver Müller, Sabine Donner, Tobias Klinder, Ralf Dragon, Ivonne Bartsch, Frank Witte, Alexander Krüger, Alexander Heisterkamp, and Bodo Rosenhahn. Model based 3d segmentation and oct image undistortion of percutaneous implants. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 6893 of *LNCS*, pages 454–462. Springer Berlin / Heidelberg, 2011.

[85] Oliver Müller, Sabine Donner, Tobias Klinder, Ivonne Bartsch, Alexander Krüger, Alexander Heisterkamp, and Bodo Rosenhahn. Compensating motion artifacts of 3d in vivo sd-oct scans. In *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, volume 7510, pages 198–205, 2012.

[86] Oliver Müller and Bodo Rosenhahn. Global consistency priors for joint part-based object tracking and image segmentation. In *Winter Conference on Applications of Computer Vision (WACV)*, pages 1129–1136, 2017.

[87] Oliver Müller, Michael Y. Yang, and Bodo Rosenhahn. Slice sampling particle belief propagation. In *IEEE International Conference on Computer Vision (ICCV)*, pages 1129–1136, 2013.

[88] Radford M. Neal. Slice sampling. *The Annals of Statistics*, 31(3):705–767, 2003.

[89] Sebastian Nowozin and Christoph H. Lampert. Structured learning and prediction in computer vision. *Foundations and Trends in Computer Graphics and Vision*, 6:185–365, 2011.

[90] Jason Pacheco and Erik Sudderth. Proteins, particles, and pseudo-max-marginals: A submodular approach. In David Blei and Francis Bach, editors, *International Conference on Machine Learning (ICML)*, pages 2200–2208. JMLR Workshop and Conference Proceedings, 2015.

[91] Jason Pacheco, Silvia Zuffi, Michael J. Black, and Erik Sudderth. Preserving modes and messages via diverse particle selection. In *International Conference on Machine Learning (ICML)*, 2014.

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

[92] Jian Peng, Tamir Hazan, David McAllester, and Raquel Urtasun. Convex max-product algorithms for continuous mrfs with applications to protein folding. In *International Conference on Machine Learning (ICML)*, 2011.

[93] Leonid Pishchulin, Micha Andriluka, Peter Gehler, and Bernt Schiele. Strong appearance and expressive spatial models for human pose estimation. In *International Conference on Computer Vision (ICCV)*, pages 3487 – 3494. IEEE, 2013.

[94] Varun Ramakrishna, Daniel Munoz, Martial Hebert , J. Andrew (Drew) Bagnell, and Yaser Ajmal Sheikh. Pose machines: Articulated pose estimation via inference machines. In *European Conference on Computer Vision (ECCV)*, 2014.

[95] Deva Ramanan. Learning to parse images of articulated bodies. In *Advances in Neural Information Processing Systems (NIPS)*, pages 1129–1136. MIT Press, 2007.

[96] Susanna Ricco, Mei Chen, Hiroshi Ishikawa, Gadi Wollstein, and Joel Schuman. Correcting motion artifacts in retinal spectral domain optical coherence tomography via image registration. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 100–107. Springer Berlin Heidelberg, 2009.

[97] Håvard Rue and Leonhard Held. *Gaussian Markov Random Fields: Theory and Applications*, volume 104 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, 2005.

[98] Mathieu Salzmann, Richard Hartley, and Pascal Fua. Convex optimization for deformable surface 3-d tracking. In *IEEE International Conference on Computer Vision (ICCV)*, 2007.

[99] Mathieu Salzmann, Julien Pilet, Slobodan Ilic, and Pascal Fua. Surface deformation models for non-rigid 3-d shape recovery. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 29(8):1481–1487, 2007.

[100] Mathieu Salzmann and Raquel Urtasun. Beyond feature points: Structured prediction for monocular non-rigid 3d reconstruction. In Andrew Fitzgibbon, Svetlana Lazebnik, Pietro Perona, Yoichi Sato, and Cordelia Schmid, editors, *European Conference on Computer Vision (ECCV)*, volume 7575 of *LNCS*, pages 245–259. Springer Berlin Heidelberg, 2012.

[101] Alexander G. Schwing, Tamir Hazan, Marc Pollefeys, and Raquel Urtasun. Distributed message passing for large scale graphical models. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1833–1840, 2011.

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

[102] Glenn Sheasby, Jonathan Warrell, Yuhang Zhang, Nigel Crook, and Philip H. S. Torr. Simultaneous human segmentation, depth and pose estimation via dual decomposition. *British Machine Vision Conference (BMVC) Workshop*, 2012.

[103] Jamie Shotton, Andrew Fitzgibbon, Mat Cook, Toby Sharp, Mark Finocchio, Richard Moore, Alex Kipman, and Andrew Blake. Real-time human pose recognition in parts from single depth images. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, CVPR '11, pages 1297–1304. IEEE Computer Society, 2011.

[104] Leonid Sigal and Michael J. Black. Measure locally, reason globally: Occlusion-sensitive articulated pose estimation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages 2041–2048, 2006.

[105] Erik B. Sudderth, Alexander T. Ihler, William T. Freeman, and Alan S. Willsky. Nonparametric belief propagation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1:605–612, 2003.

[106] Erik B. Sudderth, Alexander T. Ihler, Michael Isard, William T. Freeman, and Alan S. Willsky. Nonparametric belief propagation. *Communications of the ACM*, 53(10), 2010.

[107] Erik B. Sudderth, Michael, William T. Freeman, and Alan S. Willsky. Distributed occlusion reasoning for tracking with nonparametric belief propagation. In *Advances in Neural Information Processing Systems (NIPS)*, pages 1369–1376, 2004.

[108] Siyu Tang, Mykhaylo Andriluka, and Bernt Schiele. Detection and tracking of occluded people. *International Journal of Computer Vision (IJCV)*, 110(1):58–69, 2014.

[109] G. J. Tearney, M. E. Brezinski, B. E. Bouma, M. R. Hee, J. F. Southern, and J. G. Fujimoto. Determination of the refractive index of highly scattering human tissue by optical coherence tomography. 20(21):2258–2260, 1995.

[110] Andrey N. Tikhonov and Vasiliy Y. Arsenin. *Solutions of Ill-Posed Problems.* V. H. Winston & Sons, Washington, D.C.: John Wiley & Sons, New York,, 1977.

[111] Hoang Trinh and David McAllester. Particle-based belief propagation for structure from motion and dense stereo vision with unknown camera constraints. In *Proceedings of the International Conference on Robot Vision*, RobVis'08, pages 16–28. Springer-Verlag, 2008.

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

[112] Hoang Trinh and David McAllester. Unsupervised learning of stereo vision with monocular cues. In *Proceedings of the British Machine Vision Conference (BMVC)*, pages 72.1–72.11. BMVA Press, 2009. doi:10.5244/C.23.72.

[113] Andrew J. Viterbi. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Transactions on Information Theory*, 13(2):260–269, 1967.

[114] Karsten Vogt, Oliver Müller, and Jörn Ostermann. Facial landmark localization using robust relationship priors and approximative gibbs sampling. In *Advances in Visual Computing*, volume 9475, pages 365 – 376, 2015.

[115] Martin J. Wainwright, Tommi S. Jaakkola, and Alan S. Willsky. Tree-reweighted belief propagation algorithms and approximate ml estimation by pseudo-moment matching. In *Workshop on Artificial Intelligence and Statistics*, 2003.

[116] Martin J. Wainwright, Tommi S. Jaakkola, and Alan S. Willsky. Map estimation via agreement on trees: message-passing and linear programming. *IEEE Transactions on Information Theory*, 51:3697–3717, 2005.

[117] Martin J. Wainwright and Michael I. Jordan. Graphical models, exponential families, and variational inference. *Foundations and Trends in Machine Learning*, 1(1-2):1–305, 2008.

[118] B. Walsh. Markov chain monte carlo and gibbs sampling. Lecture Notes for EEB 581, 2004.

[119] Huayan Wang and D. Koller. Multi-level inference by relaxed dual decomposition for human pose segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2433–2440, 2011.

[120] Yang Wang and Greg Mori. Multiple tree models for occlusion and spatial constraints in human pose estimation. In *European Conference on Computer Vision (ECCV)*, 2008.

[121] Volker Westphal, Andrew M. Rollins, Sunita Radhakrishnan, and Joseph A. Izatt. Correction of geometric and refractive image distortions in optical coherence tomography applying fermat's principle. *Optics Express*, 10(9):397–404, 2002.

[122] Gerhard Winkler. *Image Analysis, Random Fields and Markov Chain Monte Carlo Methods: A Mathematical Introduction*, volume 27 of *Stochastic Modelling and Applied Probability*. Springer-Verlag New York, Inc., 2006.

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte
Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

[123] John Winn, Christopher M. Bishop, and Tommi Jaakkola. Variational message passing. *Journal of Machine Learning Research*, 6:661–694, 2005.

[124] Yi Wu, Jongwoo Lim, and Ming-Hsuan Yang. Online object tracking: A benchmark. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2013.

[125] Yi Yang and Deva Ramanan. Articulated pose estimation with flexible mixtures-of-parts. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, CVPR '11, pages 1385–1392. IEEE Computer Society, 2011.

[126] Yi Yang and Deva Ramanan. Articulated human detection with flexible mixtures of parts. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 35(12):2878–2890, 2013.

[127] Rui Yao, Qinfeng Shi, Chunhua Shen, Yanning Zhang, and A. van den Hengel. Part-based visual tracking with online latent structural learning. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2363–2370, 2013.

[128] Azadeh Yazdanpanah, Ghassan Hamarneh, Ben Smith, and Marinko Sarunic. Intra-retinal layer segmentation in optical coherence tomography using an active contour approach. In *LNCS, Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 5762, pages 649–656, 2009.

[129] Ju Hong Yoon, Du Yong Kim, and Kuk-Jin Yoon. Visual tracking via adaptive tracker selection with multiple features. In *European Conference on Computer Vision (ECCV)*, pages 28–41, 2012.

[130] S. H. Yun, G. J. Tearney, J. de Boer, and B. Bouma. Motion artifacts in optical coherence tomography with frequency-domain ranging. *Optics Express*, 12(13):2977–2998, 2004.

[131] Yun Zeng, Chaohui Wang, Yang Wang, Xianfeng Gu, Dimitris Samaras, and Nikos Paragios. Dense non-rigid surface registration using high-order graph matching. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010.

[132] Lu Zhang and Laurens van der Maaten. Preserving structure in model-free tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 36(4):756–769, 2014.

[133] Wei Zhong, Huchuan Lu, and Ming-Hsuan Yang. Robust object tracking via sparsity-based collaborative model. In *IEEE Computer Vision and Pattern Recognition (CVPR)*, pages 1838–1845, 2012.

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte
Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

[134] Silvia Zuffi, Oren Freifeld, and Michael J. Black. From pictorial structures to deformable structures. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3546–3553. IEEE, 2012.

[135] Silvia Zuffi, Javier Romero, Cordelia Schmid, and Michael J. Black. Estimating human pose with flowing puppets. In *IEEE International Conference on Computer Vision (ICCV)*, pages 3312–3319, 2013.

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte
Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte
Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

Lebenslauf

Oliver Müller
geboren am 16.06.1986
in Berlin

Beruf

Seit Apr. 2018	Softwareentwickler bei der Götting KG
Nov. 2010 – Mär. 2018	Wissenschaftlicher Mitarbeiter am Institut für Informationsverarbeitung (TNT), Leibniz Universität Hannover
Feb. 2009 – Mär. 2010	Studentische Hilfskraft beim Laboratorium für Informationstechnologie, Leibniz Universität Hannover
Okt. 2005 – Okt. 2010	Soft- und Hardwareentwickler bei der IWS Messtechnik GmbH

Ausbildung

Nov. 2010 – Sept. 2017	Doktorand am Institut für Informationsverarbeitung (TNT), Leibniz Universität Hannover
Nov. 2010	Diplom in Mathematik (Dipl.-Math.)
Okt. 2005 – Okt. 2010	Studium der “Mathematik mit Studienrichtung Informatik”, Anwendungsfach Bildverarbeitung an der Leibniz Universität Hannover
Aug. 2002 – Jul. 2005	Abitur am Niedersächsischen Internatsgymnasium, Bad Harzburg

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte
Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte
Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

Online-Buchshop für Ingenieure

VDI nachrichten

BUCHSHOP

Online-Shops

Fachliteratur und mehr - jetzt bequem online recherchieren & bestellen unter: www.vdi-nachrichten.com/Der-Shop-im-Ueberblick

Fahrzeugkonzepte für das 2. Jahrhundert Automobiltechnik
Vehicle Concepts for the 2nd Century of Automotive Technology
VDI-Berichte 1653

Täglich aktualisiert: Neuerscheinungen VDI-Schriftenreihen

Fortschritt-Berichte VDI
Weiterentwicklung der Endfunktionsprüfung von Verbrennungskraftmaschinen in der Großserienproduktion

Im Buchshop von vdi-nachrichten.com finden Ingenieure und Techniker ein speziell auf sie zugeschnittenes, umfassendes Literaturangebot.

Mit der komfortablen Schnellsuche werden Sie in den VDI-Schriftenreihen und im Verzeichnis lieferbarer Bücher unter 1.000.000 Titeln garantiert fündig.

Im Buchshop stehen für Sie bereit:

VDI-Berichte und die Reihe **Kunststofftechnik**:

Berichte nationaler und internationaler technischer Fachtagungen der VDI-Fachgliederungen

Fortschritt-Berichte VDI:

Dissertationen, Habilitationen und Forschungsberichte aus sämtlichen ingenieurwissenschaftlichen Fachrichtungen

Newsletter „Neuerscheinungen":

Kostenfreie Infos zu aktuellen Titeln der VDI-Schriftenreihen bequem per E-Mail

Autoren-Service:

Umfassende Betreuung bei der Veröffentlichung Ihrer Arbeit in der Reihe Fortschritt-Berichte VDI

Buch- und Medien-Service:

Beschaffung aller am Markt verfügbaren Zeitschriften, Zeitungen, Fortsetzungsreihen, Handbücher, Technische Regelwerke, elektronische Medien und vieles mehr – einzeln oder im Abo und mit weltweitem Lieferservice

VDI nachrichten | **BUCHSHOP** **www.vdi-nachrichten.com/Der-Shop-im-Ueberblick**

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

Die Reihen der Fortschritt-Berichte VDI:

1 Konstruktionstechnik/Maschinenelemente
2 Fertigungstechnik
3 Verfahrenstechnik
4 Bauingenieurwesen
5 Grund- und Werkstoffe/Kunststoffe
6 Energietechnik
7 Strömungstechnik
8 Mess-, Steuerungs- und Regelungstechnik
9 Elektronik/Mikro- und Nanotechnik
10 Informatik/Kommunikation
11 Schwingungstechnik
12 Verkehrstechnik/Fahrzeugtechnik
13 Fördertechnik/Logistik
14 Landtechnik/Lebensmitteltechnik
15 Umwelttechnik
16 Technik und Wirtschaft
17 Biotechnik/Medizintechnik
18 Mechanik/Bruchmechanik
19 Wärmetechnik/Kältetechnik
20 Rechnerunterstützte Verfahren (CAD, CAM, CAE CAQ, CIM ...)
21 Elektrotechnik
22 Mensch-Maschine-Systeme
23 Technische Gebäudeausrüstung

ISBN 978-3-18-3**86010**-4

https://doi.org/10.51202/9783186860101 - Generiert durch IP 216.73.216.60, am 24.01.2026, 01:49:56. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte
Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.