

Markov Random Field Based Structure Localisation of Vertebrae for 3D-Segmentation of the Spine in CT Volume Data

DIPLOMARBEIT

zur Erlangung des akademischen Grades

Diplom-Ingenieur

im Rahmen des Studiums

Medizinische Informatik

eingereicht von

David Major

Matrikelnummer 0425295

an der
Fakultät für Informatik der Technischen Universität Wien

Betreuung
Betreuer: Ao.Univ.Prof. Dipl.-Ing. Dr.techn. Eduard Gröller
Mitwirkung: Dipl.-Math. Dr.techn. Katja Bühler

Wien, 11.05.2010

(Unterschrift Verfasser)

(Unterschrift Betreuer)

Erklärung zur Verfassung der Arbeit

DAVID MAJOR
Pap Karoly Str. 1
9400 Sopron, UNGARN

Hiermit erkläre ich, dass ich diese Arbeit selbständig verfasst habe, dass ich die verwendeten Quellen und Hilfsmittel vollständig angegeben habe und dass ich die Stellen der Arbeit – einschließlich Tabellen, Karten und Abbildungen –, die anderen Werken oder dem Internet im Wortlaut oder dem Sinn nach entnommen sind, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht habe.

Wien, am 11. Mai 2010

.....

(Name)

Abstract

Medical Image Processing is a growing field in medicine and plays an important role in medical decision making. Computer-based segmentation of anatomies in data made by imaging modalities supports clinicians and speeds up their diagnosis making compared to doing it manually. Computed Tomography (CT) is an imaging modality for slice-wise three dimensional reconstruction of the human body in the form of volumetric data which is especially applicable for imaging of bony structures and so for the vertebral column. Most bony structures, such as vertebrae, are characterised by complex shape and texture appearances which turns its segmentation into a difficult task. Model-based segmentation approaches are promising techniques to cope with variations in form and texture of the anatomy of interest. This is done by incorporating information about shape and texture appearance gained from an imaging modality in a model. The model can then be applied to segment the object of interest in target data, however most of the model-based approaches need a model intialisation for a fast and reliable segmentation of the object of interest.

This thesis was motivated by novel works on fast anatomical structure localisation with Markov Random Fields (MRFs) and focuses on the sparse structure localisation of single vertebrae in CT scans for a subsequent model initialisation of more sophisticated segmentation algorithms. A MRF based model of appearance, which employs local information in regions around anatomical landmarks and geometrical information through connections between adjacent landmarks, is built on volumetric CT datasets of lumbar vertebrae. The MRF based model is built on a 6 landmark configuration in vertebra volumetric data and is additionally matched with target data. This is done by finding a best fit MRF matching by the Max-sum algorithm among feature points found by a decision tree based feature detection algorithm called probabilistic boosting tree (PBT). Anatomical landmark regions are described by vector spin-images and shape index histograms. Adjacency information is extracted by Delaunay tetrahedralisation where distances and gradient-related angles describe connections between adjacent regions. The results on single lumbar vertebra CT scans show that the MRF approach is applicable on volumetric CT datasets with an accuracy enough for supporting more sophisticated segmentation algorithms such as Active Appearance Models (AAMs).

Kurzfassung

Medizinische Bildverarbeitung gewinnt immer mehr an Bedeutung in der Medizin und spielt eine wichtige Rolle in der medizinischen Entscheidungsfindung. Computerbasierte Segmentierung von Organen in Daten von bildgebenden Modalitäten unterstützt Ärzte und ermöglicht eine schnellere Diagnose als bei manueller Segmentierung. Computertomographie (CT) ist eine der bildgebenden Modalitäten, die schnittweise eine drei-dimensionale Rekonstruktion des menschlichen Körpers erstellt, welche im Besonderen für die Bildgebung von Knochenstrukturen und so für die Wirbelsäule geeignet ist. Die meisten Knochenstrukturen, sowie auch Wirbelkörper, haben eine komplexe Form und Textur. Diese Tatsache macht die Wirbelkörper-Segmentierung zu einer schwierigen Aufgabe. Modellbasierte Segmentierungsverfahren sind vielversprechende Techniken, um Form- und Texturvariationen in einem Modell zu speichern. So ein Modell kann nachträglich verwendet werden, um Zielobjekte in unbekannten Daten zu erkennen, jedoch benötigen die meisten modellbasierten Verfahren eine Modell-Initialisierung für eine schnelle und zuverlässige Segmentierung.

Diese Diplomarbeit wurde durch neue Arbeiten im Bereich grobe Strukturerkennung von Organen mit Markov Random Fields (MRF) motiviert und verwendet das selbe Verfahren für die grobe Strukturerkennung von Wirbelkörpern in CT-Daten für die anschließende Modell-Initialisierung ausgeklügelterer Segmentierungsverfahren. Ein MRF-Modell, das lokale Informationen in Regionen um anatomisch signifikanten Landmarken und geometrische Informationen durch Verbindungen zwischen benachbarten Landmarken beinhaltet, wird mit Hilfe von volumetrischen CT-Daten von Lendenwirbelkörper gebildet. Das MRF-Modell basiert auf einer 6-Landmark-Konfiguration in Wirbelkörper-Datensätzen und wird durch das Finden einer bestgeeigneten MRF-Markierung mit dem Max-Sum-Algorithmus innerhalb von vorselektierten Merkmalspunkten auf Zielvolumensdaten angepasst, wobei die Merkmalspunkte in den Zielvolumensdaten durch den Probabilistic-Boosting-Tree (PBT) Ansatz selektiert werden. Das Aussehen der anatomischen Landmarkbereiche ist durch Vector-Spin-Images und Shape-Index-Histogramme gekennzeichnet. Nachbarschaftsinformationen zwischen Landmarken werden mit Hilfe der räumlichen Delaunay-Triangulierung gewonnen. Distanzen gemeinsam mit Gradienten-orientierten Winkeln kennzeichnen den geometrischen Aufbau des Modells. Die Resultate zeigen die Anwendbarkeit des Ansatzes auf volumetrischen CT-Daten mit einer ausreichenden Genauigkeit für die Unterstützung ausgeklügelterer Segmentierungsalgorithmen wie Active Appearance Modelle (AAMs).

Acknowledgements

First of all, I would like to thank Professor Eduard Gröller for accepting me as a master student and for reviewing my work.

This master thesis was carried out at the Virtual Reality and Visual Computing (VRVis) Research Company of Vienna. I want to thank VRVis for providing me with all the necessary facilities for my work and for supporting me at it.

I thank Katja Bühler for providing me with the topic and literature of this thesis and giving me the chance to write my thesis at VRVis. Additionally, I would like to thank her for providing me with advices and for reviewing my thesis.

Furthermore, I would like to give my thanks to Florian Schulze and Jana Hlinkova for introducing me to the topic and facilitating my work by providing me with valuable ideas.

Finally, my special thanks go to my parents and to my sister, who gave their best supporting me during my studies and my thesis writing. I would also like to express my gratitude to all of my family and to all my friends for their support.

Contents

Contents	x
List of Figures	xii
List of Tables	xiv
1 Introduction	2
1.1 Motivation	2
1.2 Problem Statement	6
1.3 Related Work	6
1.3.1 Model Initialisation for Segmentation	6
1.3.2 Markov Random Fields in Medical Image Processing	9
1.4 Approach of the Thesis	10
1.5 Thesis Overview	10
2 Medical Background	12
2.1 Vertebral Column	12
2.2 Vertebrae	13
2.3 Anatomical Variations	14
2.4 Pathological Variations	15
2.5 Imaging Modalities for the Vertebral Column	16
3 Methods	18
3.1 Overview	18
3.2 Markov Random Fields	20
3.2.1 Markov Random Field and the Gibbs Distribution	20
3.2.2 Solver Algorithms for Markov Random Fields	22
3.2.2.1 Max-sum Solver	23
3.2.2.2 Further Solver Algorithms	27
3.3 Low- and High-Level Feature Extraction	29
3.3.1 Low-Level Feature Extraction	29
3.3.2 High-Level Feature Extraction	35

3.4	Local Descriptors	37
3.4.1	Vector Spin-Image	38
3.4.2	Shape Index Histogram	41
3.5	Dissimilarity Measures	42
3.5.1	Euclidean Distance	42
3.5.2	χ^2 -Divergence	43
3.5.3	Earth Mover's Distance	43
3.5.4	Diffusion Distance	44
3.6	Geometrical Configuration	44
3.6.1	Delaunay Tetrahedralisation	45
3.6.2	Distances of Adjacencies	45
3.6.3	Gradient-Related Angles	46
4	MRF Sparse Structure Localisation of Vertebrae	48
4.1	Overview	48
4.2	Sparse MRF Model Building on Vertebrae	49
4.2.1	Training Data	49
4.2.2	Geometrical Configuration of the Sparse MRF Vertebra Model	51
4.2.3	Local Descriptor Configuration of the Sparse MRF Vertebra Model	54
4.3	Sparse MRF Vertebra Model Matching with Target Data	56
4.3.1	Overview	56
4.3.2	High-Level Feature Extraction in the Vertebra Target Data	57
4.3.3	Quality Assignment for the Sparse MRF Model Matching	58
4.3.4	Solving the MAP-MRF labelling problem with the Max-sum Solver	60
4.4	Implementation	64
5	Results	66
5.1	Evaluation Framework	66
5.2	Error Measures	67
5.3	Evaluation of the Results	68
5.4	Hardware and Performance	74
5.5	Summary	74
6	Conclusion and Future Work	76
6.1	Conclusion	76
6.2	Future Work	77
	Bibliography	80

List of Figures

1.1	Introduction to the Vertebral Column	3
2.1	Anatomy of the Vertebral Column	13
2.2	The Sixth Thoracic Vertebra and the Third Lumbar Vertebra	14
2.3	Anatomical Variations of Vertebrae	15
2.4	Imaging Modalities for the Vertebral Column	17
3.1	General Overview of the MRF Based Sparse Structure Localisation	19
3.2	Components of the MRF Based Labelling	21
3.3	Components of the Max-sum Solver	24
3.4	The Arc-consistency Algorithm of the Max-sum Solver	25
3.5	Sobel Mask Templates	30
3.6	Gradient Images of Vertebrae	31
3.7	Mathematical Definition of Curvature in 2D and 3D	32
3.8	Principal Curvature Images of Vertebrae	33
3.9	2D Integral Images and 2D Haar-like Features	34
3.10	Haar-like Features applied on Vertebrae	34
3.11	Probabilistic Boosting Tree (PBT) Learning Algorithm on a Synthetic Dataset	36
3.12	The Spin-Image Descriptor	38
3.13	The Shape Context Descriptor	39
3.14	The Vector Spin-image Descriptor and its Application to Anatomical Re- gions of Vertebra Volumes	40
3.15	Shape Index Histogram of an Anatomical Vertebra Region	42
3.16	Voronoi Diagram and Delaunay Triangulation	45
4.1	Overview of MRF Based Structure Localisation Steps of Vertebrae	49
4.2	Landmark Placement on Vertebrae	50
4.3	Diagnosis-relevant Landmarks on X-ray Vertebra Images	51
4.4	Vertebra CT Training Data	52
4.5	Geometrical Setup of the Model with Delaunay Tetrahedralisation	53
4.6	Geometrical Configuration of the Sparse MRF Vertebra Model	53
4.7	Local Descriptor Configuration of the Sparse MRF Vertebra Model	54
4.8	MRF Based Model Matching on Target Vertebra Volumes	56

4.9	Rotation Invariance around the Object Oriented y-Axis of the Vertebra . .	57
4.10	Results of the PBT Based Feature Detection Algorithm	58
5.1	Rotated Vertebra Matching	68
5.2	Sparse MRF Vertebra Model Extension by an Additional Landmark	69
5.3	Average Matching Error Results by the Leave-One-Out Test with Shape Index Histogram Descriptors	70
5.4	Average Matching Error Results by the Leave-One-Out Test with Vector Spin-image Descriptors	71
5.5	Average Matching Error Results by the Leave-One-Out Test with Shape Index Histogram Descriptors and Gradient-Based Edge Descriptors . . .	72
5.6	MRF Matching Results on Lumbar Vertebrae	73
6.1	The Matching Results of MRFs as an Initialisation of AAMs	76

List of Tables

2.1	Regions of the Vertebral Column	12
3.1	Shape Description of the Principal Curvatures	41

Chapter 1

Introduction

1.1 Motivation

The vertebral column (see figure 1.1) is a very important part of the body which holds one upright, supports movements and covers the spinal cord. It consists of complex shaped bony parts called vertebrae (see figure 1.1), which are connected through soft cushions called intervertebral discs. A vertebra generally consists of a vertebra body and transverse processes with a spinous process connected to it. The vertebra body consists of dense cortical bone which encompasses a porous matter called spongy bone.

Diseases of the vertebral column can have serious consequences such as backpain, one of the most common problems in industrialised countries; its diagnosis and treatment are crucial steps in medicine. It is stated in the health referendum of Statistics Austria in 2006/2007 [Aus07] that diseases of the vertebral column comprise the number one problem area, which means that one third of the population ages 15 and up, approximately 2.3 million people, are suffering from these types of diseases. Among these diseases are disc prolapse, foramen stenosis, spinal cord stenosis, osteoporosis, which can lead to fractures (see figure 1.1), and tumours. In order to diagnose these types of diseases, medical imaging modalities provide evidence for clinicians. Depending on the type of disease different modalities are appropriate. This means that as a first step only X-ray radiographs are generated, and if there is a demand for higher resolution, more sophisticated imaging modalities such as Computed Tomography (CT) (see figure 1.1) is applied.

CT reproduces the human body in transversal slices where the three dimensionality of the produced data comes from the construction of multiple slices. CT is an important method for visualisation of hard tissues, such as bony structures, and so it is well applicable on the vertebral column, providing high resolution volumetric data.

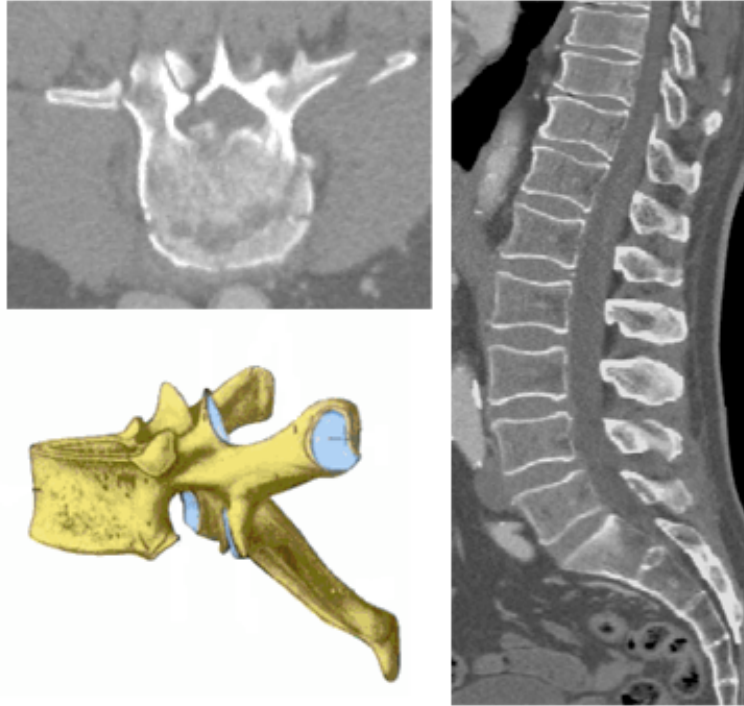


Figure 1.1: Axial CT image of a broken second lumbar vertebra (on the left top), lateral picture of the sixth thoracic vertebra [Lip06] (on the left bottom) and lateral CT image of the thoracic and lumbar part of the vertebral column (on the right).

Through the remarkable developments of medical imaging in the recent past a big demand arose for supporting clinicians with computer-aided diagnosis (CAD) which increases the accuracy and consistency of radiological diagnosis and reduces the image interpretation time. For example, computer-aided measurements of vertebra height and density can help diagnose osteoporosis. Furthermore, segmented organs, such as single vertebrae, can act as reference for registration tasks in image-guided surgery.

Segmentation of target anatomies is a difficult task due to the presence of noise in the imaging process, inhomogeneities of the tissues and the anatomical or pathological variability of target organs. Thus, different appearances due to variable densities of spongy bone of vertebrae from patient to patient, variability of thickness of intervertebral discs and diseases such as osteoporosis turn the segmentation into a difficult task. For these reasons straight-forward segmentation techniques including thresholding, as well as edge and region growing based methods are not well-suited here.

Several approaches exist for segmentation of the vertebral column in images given by different imaging modalities. Hahn and Beth [HB04] tackles the segmentation problem by searching for separation planes between vertebrae in CT data. Virtual balloons are placed and inflated within the spinal cord. After inflation, necks can be observed in the vertebra regions, and the intervertebral disc regions are characterised by swellings. The two regions can be distinguished by calculating Gaussian curvature on the balloon surface. The separation planes are finally extracted in the swelled balloon regions where the position of the plane is determined by the centre of gravity of the balloon, and the orientation of the plane is set perpendicular to the surface of the initial balloon.

In the work of Naegel [Nae07] mathematical morphology operators are applied for segmentation of vertebrae in CT data. After detecting the region of interest (ROI), which identifies the rough position of the vertebral column, a marker image is extracted out of the ROI image with the help of mathematical morphology operators, with each marker identifying a vertebra location. Finally, vertebrae are segmented with the help of the marker image and the gradient image of the ROI image by watershed-based segmentation. Labelling of vertebrae according to their anatomical names occurs by finding the twelfth thoracic vertebra (T12), which can be recognized by the fact that it is the last one from the top to the bottom that has ribs connected to it.

Model-based segmentation is an additional promising field of medical image segmentation which tackles the problem of high variability of target anatomies by modelling their shape and texture appearance. Model-based segmentation methods use prior knowledge in the form of a model for iden-

tifying objects in images. A model stands for a collection of features which describe the shape and appearance of an organ of interest. In order to detect and to segment organs of interest in volumetric data, the built model has to be matched with the previously unseen target volume data.

Klinder et al. [KWL⁺08] uses a model-based segmentation approach on CT scans and separates the problem into a global model and a local one. The global model captures the shape of the vertebral column by considering the spatial constellation of individual vertebrae, and a local model is used to enhance the segmentation by local deformations after matching the global model to the target image. The global model is matched with the target image by finding a set of optimal rigid transformations of individual vertebrae, followed by the local model matching involving local non-rigid free-form deformations.

A further model-based approach was presented [RCA06] which uses Active Appearance Models (AAMs) to segment vertebrae on X-ray radiographs. The approach uses a triple sequence of AAMs for matching from the twelfth thoracic (T12) to the fourth lumbar (L4) vertebrae, which means that the segmentation is accomplished in triplets like T12/L1/L2, L1/L2/L3 and L2/L3/L4. This approach combines the results of the fitted triplet models, where each vertebrae is fitted using the triplet sub-model in which it is central.

In order to achieve a fast and reliable segmentation with sophisticated model-based approaches such as AAMs and Active Shape Models (ASMs), a coarse initialisation of the model for matching is necessary. Donner et al. [DML⁺07], [DMLB07] copes with a sparse model initialisation for sophisticated segmentation steps such as AAMs. Donner et al. [DML⁺07], [DMLB07] uses Markov Random Fields (MRFs) for localising the sparse structure of target anatomies. MRFs model prior contextual information of local regions around interest points which are dependent on predefined adjacent interest points. The modelled information is extracted from training samples and is then compared to the same type of information extracted from a target image. The goal is to find a best fit matching of the built model with a target image based on the similarities between modelled data and target data. The best fit matching can be found with the help of a MRF solver algorithm.

The following sections provide an overview on related works of model initialisation in addition to the problem statement. MRFs in medical image processing are also presented, and a short description of the approach presented in this thesis is delivered. Finally, an overview of the thesis structure is given.

1.2 Problem Statement

Model-based segmentation methods such as ASMs and AAMs have to cope with the problem of placing the model with a considerable overlap with the object of interest on the target image at their search step. The higher the overlap of the model to the target anatomy the faster and more reliable is the searching step. For that reason many novel model-based approaches are accelerated by dividing the segmentation into a rough position localisation step followed by the model search step [DML⁺07], [DMLB07], [TBA⁺09]. An AAM-based approach was developed for vertebra segmentation so far where the model search step is very time consuming due to the large search space. A sparse structure estimation of vertebrae is crucial in order to enhance and accelerate the AAM model search by delivering initial positions where the model can be placed on. Speeding up the segmentation of individual vertebrae is of big interest because the model matching has to be accomplished on all vertebrae which are present on a given image. Furthermore, physicians could profit from the enhancement and acceleration of automatic vertebra segmentation for the diagnosis of vertebral column related diseases.

1.3 Related Work

1.3.1 Model Initialisation for Segmentation

In order to match an anatomical target with a model, at least a rough estimation of the region of interest is needed by doing it either manually or by specific approaches mentioned below. The model-based segmentation approaches described here use a rough position estimation by localising a centre point (centre of gravity) of the target organ. In the work of Ecabert et al. [EPS⁺08] a three dimensional implementation of the Generalized Hough

Transform (GHT) is applied for localisation of the heart in a fully automatic segmentation of the whole heart in three dimensional CT data. GHT encodes the description of the shape in a reference table which contains vectors pointing from the shape boundary to a reference point of the shape. The vectors are determined by the angles of gradients at boundary points according to a tangential plane at that point. At the detection step gradient orientations are measured at each edge point of a target image. Reference positions according to the extracted gradient orientations are subsequently extracted with the help of the reference table. The reference position which was selected most frequently is finally selected. Local shape variability is learned furthermore by combining the reference tables of several reference shapes. The GHT-based heart localisation suffers from high computational efforts and has to be applied after pre-processing steps such as subsampling, thresholding, smoothing and edge detection by a Sobel filter.

A further technique is used by Zheng et al. [ZBG⁺08] for heart segmentation, where probabilistic boosting trees (PBTs) are used in order to find the position and orientation of the heart. The position estimation occurs by training a classifier of a PBT given positive and negative samples based on 3D Haar wavelet features. The hypothesis whether there is an object centred at (X, Y, Z) is tested by all points of the target volume by the trained classifier, and the 100 best fits are preserved as candidate locations. After position estimation the trained position-orientation classifiers are applied in order to select the best 50 candidates out of the $100 \times N$ candidates where N is the number of hypotheses about the orientation of the heart. As the next step, trained classifiers are used to decide on the most appropriate scaling of the target anatomy model, and the outcome is ranked. The average of the top 100 candidates is the final aggregated estimate.

There are also approaches which localise not only a centre point but a sparse structure of the target organ. In the work of Roeschies and Winter [RW08] anatomical landmarks of vertebrae create an undirected graph on two dimensional axial CT slices. Gabor wavelet transformation allows the extraction of local texture information on the image and is used to generate corresponding similarities between the training and the target image.

Besides local texture information the topographic information given by the graph structure is used for matching. Furthermore, several graphs on several training images can form a bunch graph, which can improve the performance of the matching by incorporating texture and form variations in the bunch graph.

Another structure localisation approach is based on sparse MRF appearance models, which were introduced by Donner et. al [DMLB07],[DML⁺07]. In both papers a MRF framework was used for detecting the structure of hand bones in radiographs by incorporating local features and the geometrical configuration of interest points, as well as local features along adjacencies between them. Gradient Vector Flows (GVF), which detect points of high local symmetry, were used for both the description of local appearance at interest points and for the description along connections between neighbouring points. The final mapping between the MRF based model and the target interest points was made by the Max-sum algorithm [Wer07].

Novel methods present combinations of sparse object localisation techniques and more sophisticated segmentation steps on top of them. Among these multi-stage techniques is the method presented in [CCS04], where facial feature locations are approximated by a boosted cascaded classifier. The detected feature points are evaluated and filtered according to the reliabilities of the local detectors by the method called Pairwise Reinforcement of Feature Responses (PRFR). The final set of points is refined by Active Appearance Models (AAM). More than that, Tresadern et. al combine a global PCA based shape model with a MRF model based on local displacements called cascade of Combined Shape Models (c-CSM) [TBA⁺09]. The approach uses a two-stage cascade algorithm which selects obvious points by position, orientation and scale estimation by a MRF in the first step. The second step refines the result of the first step by finding a larger set of less obvious points by accepting shape variations of the global model. The approach was tested on faces and hand radiographs.

1.3.2 Markov Random Fields in Medical Image Processing

Markov Random Fields (MRFs) represented powerful tools for physicists and statisticians [Per98] before they were introduced to the image processing and computer vision community by Geman and Geman in 1984 [GG84]. MRFs model the a priori probability of image patterns that are contextually dependent. There are lots of MRF application tasks in image processing, such as image restoration [DKVC98], segmentation and texture analysis, although this section focuses on examples of medical image processing.

One of the medical applications of MRFs was developed by Held et al. [HKK⁺97], where it is applied for segmenting neighbouring tissues in brain MR images. In this approach, signal inhomogeneities are additionally modelled by an a priori MRF. The segmentation task described in this paper consists of classifying the tissue of every point as gray matter or white matter. Furthermore, MRFs are used for inhomogeneity correction of the MR images, which is important, because intensity variations can occur due to inhomogeneities of the MR magnetic fields.

Another approach incorporates MRF into the segmentation of the cardiac cavity in 2D and 3D ultrasound data. Herlin et al. [HBG⁺84] use MRFs in two types of models, a spatial model and an extended spatio-temporal model. In the spatial model the MRF is applied to segment the cavity with target data where intensity, spatial gradient information and edge information are the relevant observed features. Solving the MRF labelling problem, where the result of the labelling is whether a point is inside or outside the cavity, is done by the Iterated Conditional Mode (ICM) method. The spatio-temporal model extends the spatial one by altering the energy function according to temporal properties such as the existence of previously segmented reference images.

Besides using MRFs for segmentation purposes, there are approaches that try to capture only a sparse appearance of target objects within unknown volumes. Recently presented papers about sparse structure localisation of hand bones in radiographs using sparse MRF appearance models were already mentioned in section 1.3.1 ([DMLB07], [DML⁺07]). Tresadern et al. [TBA⁺09] (presented already in section 1.3.1) made a step forward. They ap-

plied a global PCA based shape model on a MRF based model that initialises the translation, orientation and scale of the underlying object.

1.4 Approach of the Thesis

The approach presented in this thesis was motivated by Donner et al. [DML⁺07], [DMLB07] and focuses on MRF based sparse segmentation of individual vertebrae for model initialisation of more sophisticated segmentation algorithms.

Prior information extracted from local regions around landmarks of vertebra training data and from adjacencies between them are modelled in order to capture the sparse appearance of vertebrae. Landmarks, which are manually placed in the training data, represent anatomically significant positions such as the tips of transverse processes of a vertebra, and corners of the vertebra body. Typical landmark adjacencies are connections between landmarks of the vertebra body and connections between transverse process landmarks and vertebra body landmarks. Local information from landmark regions is modelled by histogram-based local descriptors which use gradients and curvatures as low-level feature information. Adjacencies between landmarks are generated by Delaunay tetrahedralisation and additional information such as distances and angles are extracted and modelled along them. The model is additionally matched with points called feature points detected by a probabilistic boosting tree (PBT) based high-level feature extraction algorithm in target data of vertebrae where each feature point is a candidate for a certain modelled landmark. The sparse structure of vertebrae in target data is finally found by localising the modelled landmarks on a target data by finding a best fit configuration of feature points for the modelled landmarks with the help of a MRF solver algorithm. In this thesis the Max-sum solver [Wer07] was used as a MRF solver.

1.5 Thesis Overview

First of all, the anatomy of the vertebral column and medical imaging techniques are described in Chapter 2. Chapter 3 communicates the methods which are applied for the MRF based sparse structure localisation of verte-

brae. In Chapter 4 the configuration of the MRF based sparse structure localisation framework is explained. The results of the MRF based sparse structure localisation of lumbar vertebrae on CT volume data are presented in Chapter 5. Finally in Chapter 6, a conclusion of the work is provided, and possible future works are discussed.

Chapter 2

Medical Background

This section gives an introduction to the anatomy of the vertebral column and presents its possible anatomical and pathological variations. Imaging modalities of the vertebral column are also described.

2.1 Vertebral Column

The vertebral column is one of the most important parts of the body; it gives the body structure and support. On the other hand, it protects the spinal canal, which consists of bundles of nerves running from the brain to the rest of the body. The vertebral column is made up of bony parts called vertebrae. Vertebrae are stacked on each other and are connected with soft cushions called intervertebral discs, which help to absorb pressure. The vertebral column can be grouped into three main regions and two additional regions consisting of special vertebrae that are grown together (see table 2.1 and figure 2.1).

Term	Number of Vertebrae	Body Region	Abbreviation
Cervical	7	Neck	C1 - C7
Thoracic	12	Chest	T1 - T12
Lumbar	5	Low Back	L1 - L5
Sacrum	5 (fused)	Pelvis	S1 - S5
Coccyx	3-5	Tailbone	None

Table 2.1: Regions of the vertebral column

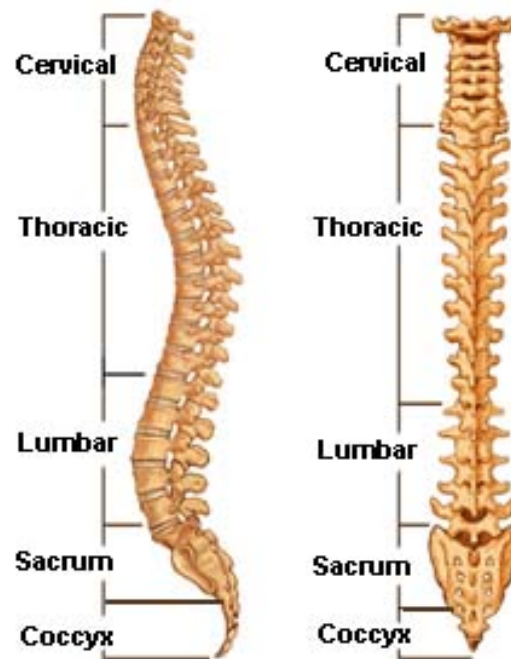


Figure 2.1: Vertebral column from the lateral direction (on the left) and vertebral column from the posterior direction (on the right), [Uni10].

Each region has a different number and different forms of vertebrae. The forms of the vertebrae depend on the function, so the cervical vertebrae are smaller than the lumbar ones. Furthermore, the vertebral column has a double S curvature, which makes it elastic, stable and provides a uniform distribution of weight. A curvature inward is called lordosis and curvatures outward represent kyphosis. According to that, the vertebral column is characterised by cervical lordosis, thoracic kyphosis, lumbar lordosis and sacral kyphosis [Lip06].

2.2 Vertebrae

Vertebrae provide protection for the spinal cord and also bear the weight put upon the vertebral column. Vertebrae have an outer shell called cortical bone that is hard and strong. The inside is made of a soft, spongy type of

bone that is called cancellous bone.

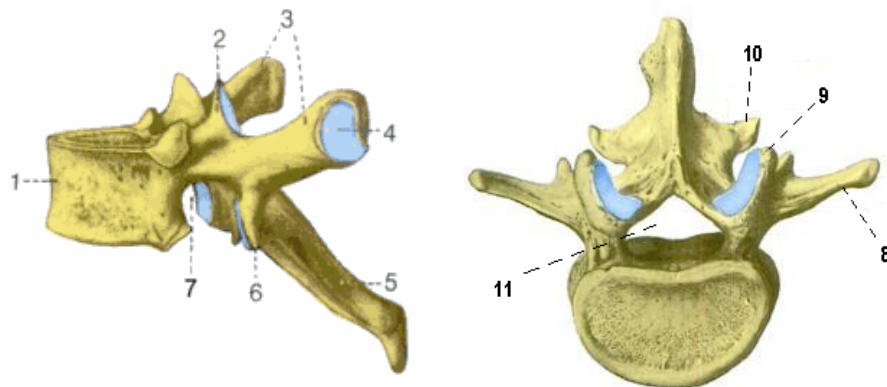


Figure 2.2: The sixth thoracic vertebra from the left lateral direction (on the left) and the third lumbar vertebra from the axial direction (on the right) [Lip06].

A vertebra consists of a large bony part called vertebra body (label 1 in figure 2.2) and a round part, which is connected to the vertebra body. Between the vertebra body and the round part is the inferior vertebral notch (label 7 in figure 2.2). The round bony part consists of the vertebral foramen (label 11 in figure 2.2), which covers the spinal canal. The spinous process (label 5 in figure 2.2) and transverse processes (label 3 and 8 in figure 2.2 where it denotes special transverse processes of lumbar vertebrae called costal processes) are further extensions of the round bony part. The ribs attach to the vertebrae at the transverse processes by the fovea costalis (label 4 in figure 2.2). Furthermore, two more types of processes exist, the superior articular processes (label 2 and 9 in figure 2.2) and the inferior articular processes (label 6 and 10 in figure 2.2), [Lip06].

2.3 Anatomical Variations

Sifting through the vertebral column from the top to the bottom, anatomical variations (see figure 2.3) can be observed as the shapes of vertebrae change, and the processes and vertebra bodies become larger (which is partly due to the fact that the lower parts have to bear more weight than the upper ones).

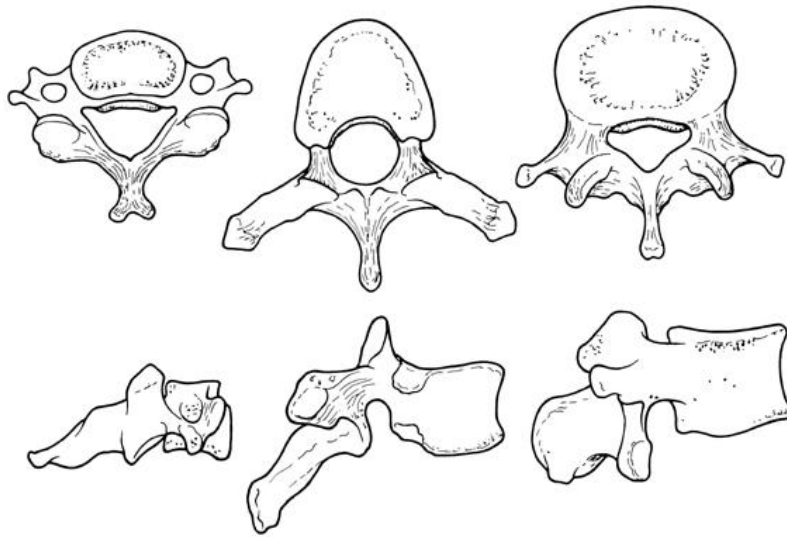


Figure 2.3: Cervical (on the left), thoracic (in the middle) and lumbar vertebrae (on the right) from axial (top row) and lateral directions (bottom row), [Min10].

In addition to that, irregular variations of the shape of single vertebrae are also possible. Examples of such irregular variations are wedge-shaped vertebrae, which can be due to hypoplasia of the vertebral body, and vertebra plana, which stands for plate-like formed vertebrae with a normal vertebral arch range. More than that, vertebra bodies with concave anterior edges and fish-formed vertebrae can be also mentioned among the irregular shape variations of vertebrae. Another occurrence of anatomical variation is fused vertebrae, which is possible either congenitally or can evolve over years and is characterised by partial or total loss of intervertebral space [RKD06].

2.4 Pathological Variations

Besides anatomical variations scoliosis, osteoporosis, spinal stenosis, fractures, osteophytes and tumours are well-known disease-related variations of the vertebral column.

Scoliosis is described as the abnormal curvature of the vertebral column in the frontal plane, which causes rotation and deformation of single vertebrae. The vertebral column deforms into a C- or S-like shape in the anterior

plane.

Osteoporosis is a bone disease in which the mineral density of the bones is reduced. It can lead to an increased risk of bone fractures. The vertebral column is one part of the body that has a high risk of being affected by osteoporosis. Bone spurs may appear on vertebrae, or their size may shrink, causing vertebral bodies to lose their height.

Osteophytes, tumours, fractures and spinal stenosis cause shape variations at certain locations of the vertebral column, so these have effect on single vertebrae. Osteophytes are bony spurs that can grow on any part of the vertebrae. Spinal stenosis stands for narrowing of the vertebral foramen in the case of degenerative changes. Tumours and fractures refer to unpredictable changes of the vertebra shape [RKD06].

2.5 Imaging Modalities for the Vertebral Column

The basic imaging modality for the vertebral column is X-ray radiographs, which are projected from two directions (anterior or posterior side and lateral side). X-ray radiographs are generated by casting X-rays of an X-ray beam towards the anatomy of interest and by measuring the absorption of X-rays based on different densities and composition of tissues (see figure 2.4 on the left). Furthermore, 3D imaging modalities such as Computed Tomography (CT) and Magnetic Resonance Imaging (MRI) are applied in order to support diagnosis in the case of special diseases like tumours, fractures or intervertebral disc degenerations and to overcome occlusions by other organs. CT is an X-ray based method for the generation of images of transversal slices of the underlying anatomy. Three dimensionality of CT is then the outcome of the sum of the generated transversal slice images (see figure 2.4 on the right). MRI is an imaging technique based on the magnetisation of hydrogen atoms in the body. The alignment of the hydrogen nuclei is changed by magnetising them by the MRI scanner which is defined as their initial position. The alignment of the magnetisation can be then transformed into a rotating field by additional radio frequency fields. By switching off the radio frequency fields, the relaxation times to the initial position typical for different tissues can be measured according to different contrasts on the con-

structured image [RKD06] (see figure 2.4 in the middle).

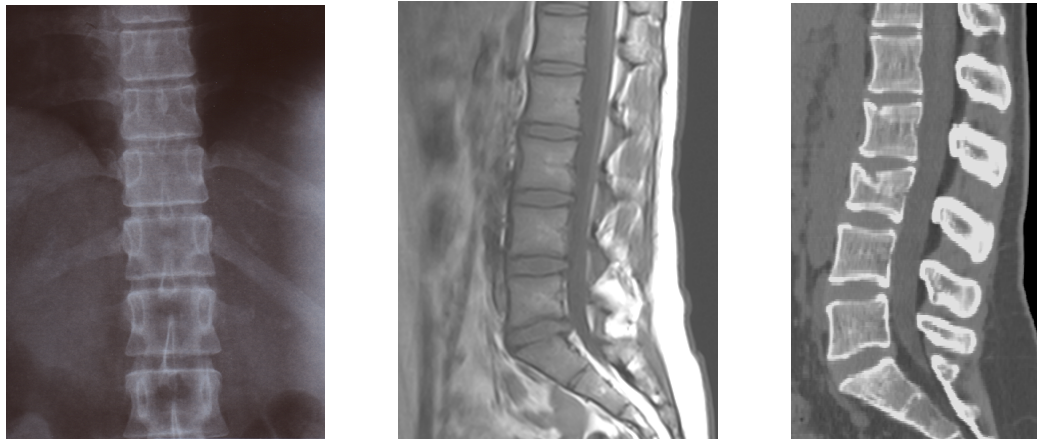


Figure 2.4: Coronal radiograph of the thoracic vertebral column, sagittal MRI and CT of the lumbar vertebral column (from left to right).

CT and MRI are modalities that produce gray level based images. Whereas CT is better suited for imaging of bony parts of the body, MRI is mainly used for the diagnosis of soft tissues of the body. For CT data the resolution of axial slices generally lies between 0.25 and 2 millimetres and the distance between slices use to lie in a range from 0.2 to 3 millimetres. MRI uses arbitrary directions for imaging of the vertebral column, sagittal and coronal slices are used.

Multidetector-Spiral-Computertomographs are recommended because of several advantages over Single-Slice-Computertomographs in the case of using CT for the diagnosis of the vertebral column. Spiral-CTs reproduce the volume with a much higher longitudinal resolution and in shorter periods than Single-Slice-CTs. CT images are useful for diagnosis of fractures of the vertebral column, tumours, metastases and canal compressions (i.e. slipped discs). In case of analysing the vertebral column by MRI, sagittal T1- and T2-weighted images are reconstructed where T1 and T2 denote two different types of relaxation times. The T1-weighted images are used for the diagnosis of bone marrow, and T2-weighted ones are used to evaluate the myelon, the spinal canal and the discs [RKD06], [HL07].

Chapter 3

Methods

3.1 Overview

Markov Random Field (MRF) based sparse structure localisation applied to medical data finds an anatomy of interest by labelling its anatomically significant points called landmarks in unknown data called target data. It can be separated into two major parts, MRF based model building and model matching (see figure 3.1).

MRF based model building - At the MRF based model building, the sparse appearance and shape of an anatomy of interest is modelled. The model represents a graph where its nodes capture the appearance of regions around landmarks, and its connections define adjacencies between landmarks establishing a geometrical configuration. The nodes of the model are called objects and the connections compatibilities in this thesis. Appearance information is prepared in the form of local descriptors, and the shape is reflected by the geometrical configuration which involves distance and angle information. The model objects are characterised by local descriptors which are calculated as the average of local descriptors around manually placed corresponding landmarks in a set of data called training data. Distances and angles are extracted from a corresponding set of adjacent landmarks of training data where landmark adjacencies are defined by Delaunay tetrahedralisation. Gaussians of distances and angles along adjacencies setup a geometrical configuration and describe the compatibilities of the model.

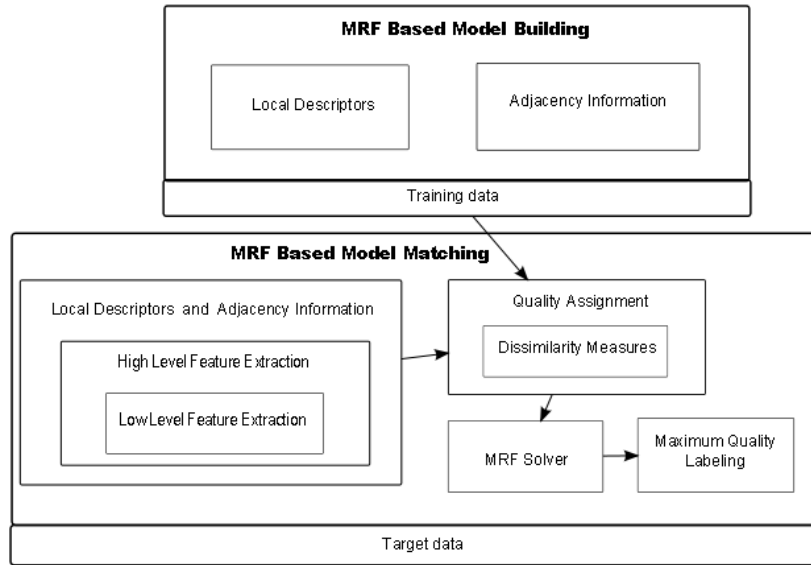


Figure 3.1: MRF Based Structure Localisation involves a high variety of methods which are visualised by rectangles enclosed in the two major parts of the framework: MRF based model building and model matching.

MRF based model matching - As a first step, information delivered by local descriptors is extracted around points with features of interest called feature points which are detected automatically by a high-level feature extraction algorithm on target data. In addition, information along connections between the feature points are extracted according to the geometrical configuration of the model. These two types of information are compared to those of the built model yielding qualities which describe how well feature points match the model objects and how well the connections between feature points called edges fit to the compatibilities of the model. The feature points are called labels in the context of the MRF based model matching because each of them are possible candidates for labelling the landmarks of the target anatomy. The goal of the MRF based model matching is to find the label and edge configuration with the highest overall quality. The label-edge configuration called labelling with maximal overall quality is finally found by a MRF solver.

All of the methods essential for the MRF based model building and matching are described in the following sections of this chapter.

3.2 Markov Random Fields

Markov Random Fields (MRF) are described first as it is the main technique used in this thesis and its concepts are important for both the MRF based model building and matching. Thus, MRFs model the qualities resulting from model matchings with target data. Of interest is a model match with maximum quality. For finding a maximum quality matching, MRF solvers are applied. A theoretical explanation of MRFs is delivered and MRF solvers are presented in the following.

3.2.1 Markov Random Field and the Gibbs Distribution

MRFs [Sta95] represent graphs $M = (T, E)$ where T are the nodes and E represents adjacencies of the graph. Two nodes represented by objects are considered as neighbours in presence of an adjacency represented by a compatibility $e \in E$. Neighbours of an object $t \in T$ are denoted by t' . A finite set of discrete random variables X builds a random field where each variable X_t is assigned to a node $t \in T$. A realisation $x_t \in \mathcal{L}_t$ called label is assigned to each of them by $X_t = x_t$ where \mathcal{L}_t corresponds to a discrete set of labels out of the $|T|$ -tuple of labels $\mathbb{X} = \mathcal{L}^{|T|} = \mathcal{L}_1 \times \dots \times \mathcal{L}_T$. Connections between labels of adjacent objects are called edges $a_e \in A_e$ where adjacencies are defined by compatibilities $e \in E$. The joint probability $P(X = \mathbf{x}) = P(X_1 = x_1, \dots, X_T = x_T)$ defines a configuration where every random variable has a realisation (see figure 3.2).

X describes a MRF if the local characteristic of a probability measure depends only on the knowledge of neighbouring nodes and the probability measure is positive:

$$\begin{aligned} P(\mathbf{x}) &> 0, \forall \mathbf{x} \in \mathbb{X} \\ P(x_t | x_{T-t}) &= P(x_t | x_{t'}). \end{aligned} \tag{3.1}$$

Formula (3.1) depicts the Markov-property of the random field, which means that only neighbouring labels have direct interaction with each other.

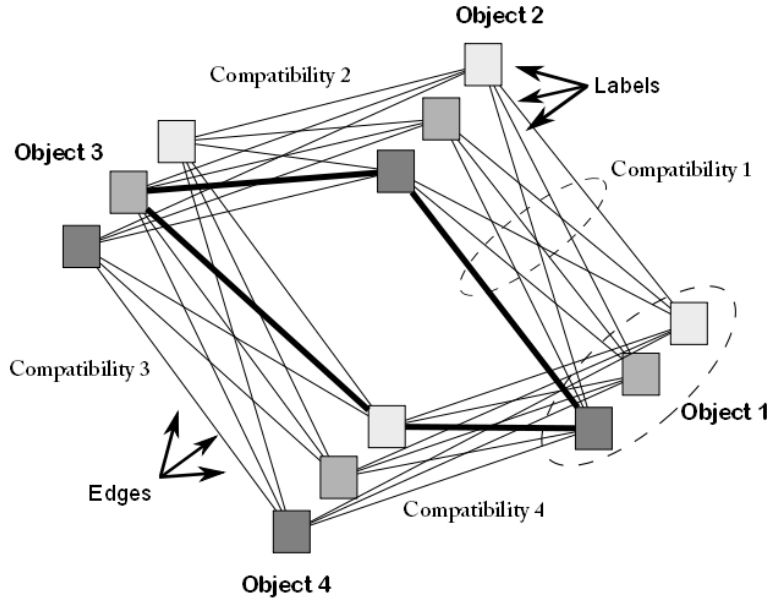


Figure 3.2: The MRF graph includes objects $t \in T$ and labels. The objects are connected by compatibilities $e \in E$ where each compatibility contains edges $a_e \in A_e$ connecting labels. A labelling $X = \mathbf{x}$ is denoted by thick edges which assigns a label x_t to each object $t \in T$ [DMLB07].

A simple example for MRF application is segmentation of images with dark and bright regions. In this case, intensities of dark and bright regions and intensity differences between the two types of regions could be modelled by a MRF graph. The intensity information would be modelled by objects and the intensity difference between them according to a predefined neighbourhood system (e.g. 4-point-neighbourhood) would provide compatibility information of the MRF graph. Applying the graph on target images, each image point gets a label, dark or bright, assigned under consideration of the modelled intensities and intensity differences along the defined neighbourhood system.

For establishing a simple way to specify the joint probabilities $P(X = \mathbf{x})$, the Hammersley-Clifford theorem states that an equivalence can be established between the local properties of Markov Random Fields and global

ones of Gibbs distributions. That way joint probabilities can be represented by potential functions assigned to sub-configurations of \mathbf{x} where appropriate potential functions can be chosen for desired system behaviours [Sta95]. The equivalence criterion between MRFs and Gibbs distributions can be explained as follows: An energy $U(\mathbf{x})$ on labellings $\mathbf{x} \in \mathbb{X}$ is defined as

$$U(\mathbf{x}) = \sum_{c \in C} V_c(\mathbf{x}) \quad (3.2)$$

where a clique $c \in C$ describes a sub-graph of T where every object $t \in T$ is connected to all the other ones and $V_c(\mathbf{x})$ corresponds to the clique potential which is the prior probability assigned to a sub-configuration \mathbf{x}_c . Examples for cliques are single objects, pairs of objects and triples of objects. The Gibbs measure is defined as

$$P(\mathbf{x}) = \frac{e^{-U(\mathbf{x})}}{Z} \quad (3.3)$$

where

$$Z = \sum_{\mathbf{x} \in \mathbb{X}} e^{-U(\mathbf{x})} \quad (3.4)$$

is the partition function. In order to make the relation between Markov Random Fields and Gibbs distributions, nearest neighbour Gibbs potentials have to be defined. By considering arbitrary subsets $s \in S$ of T , a potential is a nearest neighbour Gibbs potential if $V_s(\mathbf{x}) = 0$ only when s is not a clique, which means that only clique potentials contribute to the energy. Therefore, it is visible that the nearest neighbour Gibbs measure induced by the nearest neighbour Gibbs potential builds the necessary equivalence criterion between Gibbs distributions and Markov Random Fields [KS80].

3.2.2 Solver Algorithms for Markov Random Fields

Of interest is to find a configuration of the MRF which is the most similar to the priors of a presented model, thus, having a maximal joint distribution. In this section solver algorithms are introduced for Markov Random Fields that compute the maximum posteriori estimate (MAP) of the MRF with discrete variables by

$$\mathbf{x}^* = \arg \max_{\mathbf{x} \in \mathbb{X}} P(\mathbf{x}) \quad (3.5)$$

which is equivalent to finding a configuration of the nearest neighbour Gibbs distribution with maximal probability [DMLB07]. In equation (3.5) x^* denotes an optimal labelling delivered by the MAP estimate. In the case of the dark-bright segmentation example of section 3.2.1, the segmentation of dark-bright regions on a target image can be found by calculation of the MAP estimate out of all possible labelling probabilities. First, the Max-sum solver is presented, which finds a solution for the labelling problem by maximising a sum of bivariate functions of discrete variables. Two other solver algorithms are explained additionally.

3.2.2.1 Max-sum Solver

The Max-sum solver of Schlesinger originally proposed in 1976 [Sch76] is reviewed by Werner [Wer07] where the (binary) Max-sum problem can be defined as maximising a sum of unary and binary functions of discrete variables. Werner [Wer07] describes a linear programming relaxation of the Max-sum problem, however, the Max-sum problem is solved with the help of the dual of the linear program as doing that way is more efficient. Furthermore, the final solution is found by formulating a Constraint Satisfaction Problem (CSP) as the solution has to satisfy unary and binary constraints. This section provides an overview of the Max-sum problem formulation, describes the CSP and arc-consistency as a solver for CSPs and explains the steps for solving the Max-sum problem.

Max-sum problem formulation - As already mentioned before the Max-sum labelling problem is defined as maximising the unary and binary function of discrete variables, which is a general NP-hard optimisation problem [Wer07]. If applied for MAP inference of MRFs the unary and binary functions correspond to joint probabilities $P(X = x)$. Probabilities P are represented here by qualities Q which are obtained through comparison of modelled descriptors to target-extracted ones by dissimilarity measures described in section 3.5. That way, qualities describe object-to-label (unary) and compatibility-to-edge (binary) relationships. The quality of a labelling $Q(X = x)$ can be

formulated as

$$Q(\mathbf{x}|\mathbf{q}) = \sum_{t \in T} q_t(x_t) + \sum_{(t,t') \in E} q_{tt'}(x_t, x_{t'}). \quad (3.6)$$

where each object $t \in T$ is assigned a label $x_t \in \mathcal{L}_t$, and numbers $q_t(x_t)$, $q_{tt'}(x_t, x_{t'}) \in \mathbb{R} \cup (-\infty)$. $q_t(x_t)$ are qualities for the label x_t of the object t and $q_{tt'}(x_t, x_{t'})$ are qualities for an edge between the object t with label x_t and the object t' with the label $x_{t'}$ related to a labelling $\mathbf{x} \in \mathbb{X}$. Labels not necessarily corresponding to a labelling \mathbf{x} are denoted by $\hat{x} \in \mathcal{L}$. Furthermore, a pencil (t, t', \hat{x}) is defined as the set of edges from a node (t, \hat{x}) to all nodes of a neighbouring object t' (see figure 3.3).

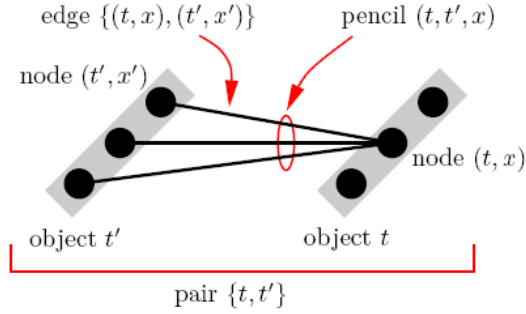


Figure 3.3: Description of the components of the MRF graph for the Max-sum solver algorithm [Wer07].

Finding a solution to the labelling problem means maximising the function (3.6) in order to get an optimal labelling:

$$L_{M,\mathcal{L}}(\mathbf{q}) = \arg \max_{\mathbf{x} \in \mathbb{X}} Q(\mathbf{x}|\mathbf{q}). \quad (3.7)$$

Constraint Satisfaction Problem (CSP) - CSP is explained first as it has to be formulated and solved in order to find the final solution to the Max-sum problem.

The CSP is defined as finding a labelling which satisfies given unary and binary constraints. Thus, a labelling is of interest which passes through some or all of the given nodes and edges. A CSP can be formulated as

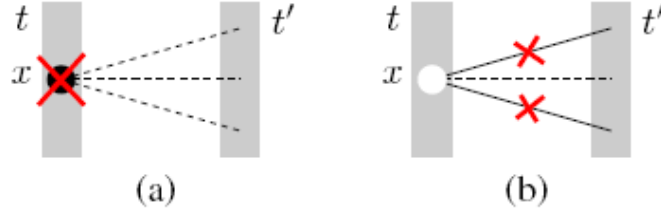


Figure 3.4: „The arc-consistency algorithm deletes (a) nodes not linked with some neighbour by any edge, and (b) edges lacking an end node. “[Wer07]

$$\bar{L}_{M,\mathcal{L}}(\bar{\mathbf{q}}) = \left\{ \mathbf{x} \in \mathbb{X} \mid \bigwedge_t \bar{q}_t(x_t) \wedge \bigwedge_{\{t,t'\}} \bar{q}_{tt'}(x_t, x_{t'}) = 1 \right\} \quad (3.8)$$

where $\bar{q}_t(\hat{x}), \bar{q}_{tt'}(\hat{x}, \hat{x}') \in \{0, 1\}$ are set to 1 or 0 if a node respectively edge satisfies a given constraint or not (in the case of Max-sum it is the constraint that node and edge qualities have to be maximal). A CSP is satisfiable if $\bar{L}_{M,\mathcal{L}}(\bar{\mathbf{q}}) \neq \emptyset$. A satisfiable CSP can be found by terms of local constraints called arc-consistency (see figure 3.4). A CSP is arc-consistent if

$$\bigvee_{\hat{x}'} \bar{q}_{tt'}(\hat{x}, \hat{x}') = \bar{q}_t(\hat{x}) \quad (3.9)$$

with $\{t, t'\} \in E$ and $\hat{x} \in \mathcal{L}$. Checking a CSP for satisfiability can be done with an arc-consistency algorithm where a solution to the CSP found by the arc-consistency algorithm is called the kernel of the CSP [Wer05], [Wer07].

Solving the Max-sum problem - In general, the Max-sum problem can be solved by linear programming relaxations. A relaxed labelling is defined as vector ϱ with the components $\varrho_t(\hat{x})$ and $\varrho_{tt'}(\hat{x}, \hat{x}')$ representing numbers

assigned to labels and edges satisfying

$$\begin{aligned} \sum_{\hat{x}'} \varrho_{tt'}(\hat{x}, \hat{x}') &= \varrho_t(\hat{x}), \quad \{t, t'\} \in E, \hat{x}, \hat{x}' \in \mathcal{L} \\ \sum_{\hat{x}} \varrho_t(\hat{x}) &= 1, \quad t \in T \\ \varrho &\geq 0. \end{aligned} \tag{3.10}$$

In the previous sections labellings $\mathbf{x} \in \mathbb{X}$ were mentioned where each object $t \in T$ gets exactly one label assigned $x_t \in \mathcal{L}_t$. The formula (3.10) presents an alternative representation where objects are allowed to be undecided which means that multiple labels can be assigned to an object with different weights. Thus, if ϱ in (3.10) denotes a binary vector it represents a decided labelling and ϱ is said to be undecided if it contains non-integer elements. The set ϱ satisfying the conditions in (3.10) is denoted by $\Lambda_{M, \mathcal{L}}$. Thus, the relaxed Max-sum problem can be described as the linear program

$$\Lambda_{M, \mathcal{L}}(\mathbf{q}) = \arg \max_{\varrho \in \Lambda_{M, \mathcal{L}}} \langle \mathbf{q}, \varrho \rangle \tag{3.11}$$

In (3.11) the quality of a relaxed labelling is the scalar product of \mathbf{q} and ϱ indicated by $\langle \mathbf{q}, \varrho \rangle$. $\Lambda_{M, \mathcal{L}}(\mathbf{q})$ coincides with $L_{M, \mathcal{L}}(\mathbf{q})$ if it represents a decided labelling and delivers the solution to the labelling problem formed by labels which have values assigned unequal to 0 after the relaxation.

It is inefficient to compute the linear programming relaxation for large data sets as it is the case in most of the computer vision tasks. The linear programming dual of the Max-sum problem can be used for solving the problem more efficiently by the minimisation

$$\mathbf{U}^*(\mathbf{q}) = \min_{q' \sim q} \mathbf{U}(\mathbf{q}') \tag{3.12}$$

of the upper bound

$$\mathbf{U}(\mathbf{q}) = \sum_t \max_{\hat{x}} q_t(\hat{x}) + \sum_{\{t, t'\}} \max_{\hat{x}, \hat{x}'} q_{tt'}(\hat{x}, \hat{x}') \tag{3.13}$$

to the Max-sum problem. The minimisation of the upper bound is established by equivalent transformations \mathbf{q}' . An equivalent transformation can be

formulated as adding a number μ to $q_t(\hat{x})$ and subtracting the same number from all edges in the pencil (t, t', \hat{x}) . Equivalent transformations are redefinitions of the problems by leaving $Q(x|q)$ unchanged. However, the upper bound $U(q)$ is not invariant to equivalent transformations which leads to the minimisation of the upper bound by equivalent transformations in order to approach a labelling with maximum quality [Wer07].

Finally, the result to the Max-sum problem is delivered by a satisfiable CSP on labels and edges with maximal quality delivered by the approximated minimal upper bound $U^*(q)$. It is remarked by Werner [Wer05], [Wer07], that a non-empty kernel of the CSP formed by the labels and edges with maximal quality is necessary but not sufficient for minimal upper bound of the problem, which means that the algorithm sometimes might not find the maximum quality MRF labelling.

3.2.2.2 Further Solver Algorithms

Alternative and novel methods for minimisation of the Markov Random Field energy have been introduced recently. One of these new methods, called Iterated Cross Entropy with Partition Strategy (ICEPS), is described in this section [WC07]. Besides new approaches a well-known method, called Simulated Annealing, is explained in the application for solving the MAP-MRF labelling problem. Both methods are based on the maximisation of the MAP probability as expressed in (3.5), which can be formulated as the minimisation of the energy function U of the underlying Gibbs distribution [WC07] by

$$x^* = \arg \min_{x \in \mathbb{X}} U(x) \quad (3.14)$$

where $U(\cdot)$ is the energy function from (3.2) and x^* is the optimal configuration of labels.

Cross Entropy Energy Minimisation - A novel energy minimisation method was introduced by Wu and Chung [WC07], which is called Iterated Cross Entropy with Partition Strategy (ICEPS). This method associates an estimation problem with the optimisation problem by estimating probabilities for configurations with low energies. The algorithm calculates the estimates for an optimal configuration based on formula (3.14). It estimates the probability P

so that the energy of a configuration remains within a threshold $\vartheta \in \mathbb{R}$ with the following formula

$$P(U(X) \leq \vartheta) = \sum_{\mathbf{x}} I_{\{U(\mathbf{x}) \leq \vartheta\}} p(\mathbf{x}; \mathbf{v}) \quad (3.15)$$

where $I_{\{event\}}$ is an indicator function which is 1 if the event is true and 0 otherwise. $p(\cdot; \mathbf{v})$ is the probability density function for configurations $\mathbf{x} \in \mathbb{X}$ where \mathbf{v} is its parameter vector which indicates possible configurations by assigning high probabilities to preferable labels. The formula (3.15) estimates the probability for the situation where the energy of a configuration $\mathbf{x} \in \mathbb{X}$ is less than a threshold ϑ . The optimisation problem is solved with a multi-level optimisation approach where sequences of levels $\vartheta_1, \vartheta_2, \dots, \vartheta_L$ and parameter vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_L$ are generated so that ϑ_L is close to the optimal energy $\vartheta^* = \min_{\mathbf{x} \in \mathbb{X}} U(\mathbf{x})$, and \mathbf{v}_L corresponds to a parameter vector which assigns high probability density to configurations that correspond to low energies. Thus, the ICEPS algorithm focuses on high performance configurations with energy lower than a threshold ϑ and calculates the occurrence of labellings out of the high performance configuration set. This statistical information is used to guide subsequent new configuration set generating iterations of the algorithm. This way, the Cross Entropy between the optimal importance sampling density and the target density can be minimised [WC07]. Cross Entropy is embodied by the Kullback-Leibler distance, and the optimal importance sampling density is calculated by focusing on important configuration subsets with low energies.

Simulated Annealing - The well known heuristic technique Simulated Annealing is used in the context of Markov Random Fields for finding the MAP estimate of the labelling problem by minimisation of the energy of the system. In the work of Mohapatra et al. [MNP06] a labelling problem is solved by the application of a MRF model for color image segmentation. The goal is to find an optimal pixel labelling for the segmentation of the observed image where the observed image is assumed to be a degraded version of the true labels by considering a Gaussian distributed degradation.

Simulated Annealing heats the system with a high temperature F_{in} and causes permanent internal energy changes by perturbing the system with

Gaussian disturbance. The energy U' of the perturbed system is calculated and $\Delta U = U - U'$ is analyzed. If $\Delta U > 0$ then the perturbed system is accepted as a new configuration with a probability of 1, otherwise the new configuration is accepted with a probability $\exp(\Delta U/F)$. The temperature F has to be decreased in every pass until it decreases to a sufficiently low value. The energy at this low temperature delivers the MAP estimate and so produces the final labelling of the system.

3.3 Low- and High-Level Feature Extraction

The low- and high-level feature extraction is an important preparation step for the MRF based model matching. In order to enhance the matching, the space of possible candidate points within target data has to be reduced to points with given characteristics called features. At first, low-level features, such as Haar-like features [VJ02], gradient magnitudes and curvatures are extracted. High-level features are learned based on the extracted low-level ones at predefined points of training data by decision tree based classification and can be detected on a target data. Points detected by the high-level feature extraction algorithm on target data are the feature points. The feature points detected on a target data are finally the labels of the MRF based model matching.

3.3.1 Low-Level Feature Extraction

Feature extraction techniques provide interpretation of images at sampled image locations or predefined regions. Images can be defined in either two or three dimensions. A two-dimensional (2D) real world image is considered to be a function of two real variables, for example, $I(x, y)$ with I as the intensity of the image at the real coordinate point (x, y) . Intensities of medical images can be represented by gray scales. A three-dimensional (3D) image called volume is defined as the function $I(x, y, z)$ at the real coordinate point (x, y, z) . 2D and 3D images are divided into elements called pixels in the 2D case and voxels in the 3D case. Real valued points (x, y) and (x, y, z) are mapped to certain pixels and voxels depending on the pixel- and voxel-sizes of the image.

There are many motivations for using features based on different regions instead of only intensity information at each image point. Features represent ad-hoc domain knowledge, and systems based on extracted features operate faster than point based ones as stated by Viola and Jones [VJ02]. The classifiers of the high-level feature extraction algorithm can be trained based on multiple low-level features. In the first step low-level features like intensity, gradient and curvature are extracted and processed locally without any additional information like shape or spatial integration. These features can be incorporated into higher level features where information about shape and spatial integration are also used [NA08]. This section communicates the low-level feature extraction methods.

Image Gradients - Gradients are first order differential quantities which measure how intensities vary in predefined regions around a point and give the steepest ascent at that point. The gradient of an image I at the point (x, y, z) is defined as the three-dimensional vector $g = \nabla I = [I_x, I_y, I_z] = [\frac{\partial I}{\partial x}, \frac{\partial I}{\partial y}, \frac{\partial I}{\partial z}]$ where I_x, I_y, I_z are the partial derivatives of I at the point (x, y, z) .

1	0	-1
2	0	-2
1	0	-1

1	2	1
0	0	0
-1	-2	-1

Figure 3.5: 2D Sobel mask templates for m_x (on the left), and for m_y (on the right) [NA08].

One of the well-known approximation approaches of gradients is computing masks which are convolved with the underlying volume in order to estimate the first derivatives at each position of the volume. The Sobel operator (see figure 3.5) is an operator which consists of three masks m_x, m_y, m_z where each of them corresponds to one of the dimensions x, y, z of the volume in order to approximate the partial derivatives along each dimension. The three partial derivatives I_x, I_y, I_z which set up the gradient vectors in each location of the volume can be calculated by convolution of the volume with the three

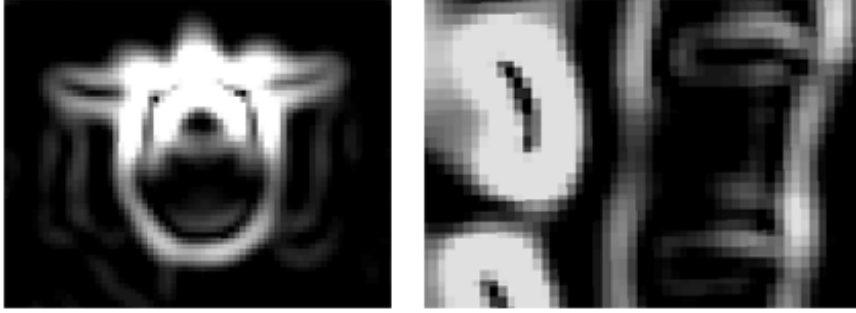


Figure 3.6: Axial gradient image (on the left) and lateral gradient image of a vertebra (on the right). Bright to dark intensities encode high to low gradient magnitudes.

masks as follows:

$$\begin{aligned} I_x &= I \star m_x \\ I_y &= I \star m_y \\ I_z &= I \star m_z. \end{aligned} \tag{3.16}$$

The gradient volume is the result of the application of all three masks on the intensity function I of the volume where the three gradient components I_x, I_y, I_z are assigned to each point of the volume. The gradient magnitude is defined by $G = \sqrt{I_x^2 + I_y^2 + I_z^2}$ as the rate of change at a given point [NA08]. Points with high G are possible raw estimates for edge points where “high” can be realised as higher than a given threshold. Gradient images of vertebrae are visible in figure 3.6.

Curvatures - Curvature can be defined mathematically as the rate of change of the direction of a curve. Thus, considering a point on the curve, then the curvature is the rate of change of angle $\partial\omega$ with respect to arc length ∂S which can be formulated as $\kappa = \frac{\partial\omega}{\partial S} = \frac{2\pi}{2\pi R} = \frac{1}{R}$ after Roberts [Rob01], where R is called radius of curvature, which is the radius of the circle having the greatest contact with the curve, called oscillating circle (see figure 3.7 on the left). The concept of curvature can be extended into three dimensions by intersecting a surface with planes. At the curve resulting from the intersection, the curvature can be calculated where the most useful ones are those de-

fined by planes orthogonal to the surface giving the maximum and minimum curvatures (principal curvatures) [Rob01].

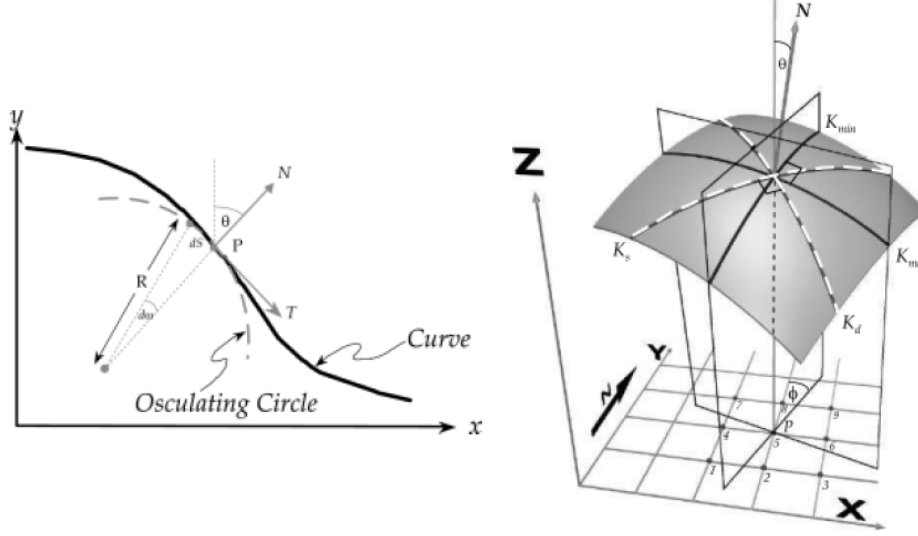


Figure 3.7: Mathematical definition of curvature where the curvature for a particular point P on a curve is defined as the reciprocal of the radius R of the oscillating circle (on the left); picture of the min- and max-curvatures in three dimensions which are defined at a point P by the intersection of planes orthogonal to the surface (on the right) [Rob01].

Curvatures are defined as second derivatives of a function. If given a Hessian matrix H by

$$H = \begin{bmatrix} I_{xx} & I_{xy} & I_{xz} \\ I_{yx} & I_{yy} & I_{yz} \\ I_{zx} & I_{zy} & I_{zz} \end{bmatrix} \quad (3.17)$$

with $I_{xx} = \frac{\partial^2 I}{\partial x^2}$, $I_{xy} = \frac{\partial^2 I}{\partial x \partial y}$, ..., $I_{zz} = \frac{\partial^2 I}{\partial z^2}$ second-order partial derivatives of the image function $I(x, y, z)$, then the largest eigenvalue λ_1 with the eigenvector e_1 of the matrix H corresponds to the first principal curvature κ_1 and the second largest eigenvalue λ_2 with eigenvector e_2 of H denotes the second principal curvature κ_2 of an underlying shape. The eigenvectors e_1 and e_2 correspond to orthogonal directions of the two curves with minimal and maximal curvature. Images of the two principal curvatures on vertebra volumes are visible in the figure 3.8.



Figure 3.8: κ_1 -curvatures of a vertebra volume on the left and κ_2 -curvatures on the right. Bright to dark intensities encode high to low principal curvatures.

In the work of Kindlmann et al. [KWTM03], an efficient approach is presented for estimation of the two principal curvatures for each point in volumetric data. This approach uses the fact that surface curvatures are defined by the relationship between small positional changes on the surface and the resulting changes of surface normals. It uses the gradient vectors \mathbf{g} for approximation of the surface normals and the Hessian matrix H for determining how the gradient \mathbf{g} changes as a function of infinitesimal changes of position in \mathbb{R}^3 . The gradient can change in length and in direction, however for describing curvatures only the gradient direction changes are of interest. Thus, the principal curvatures κ_1 and κ_2 are extracted from a matrix which isolates the gradient direction changes from the Hessian matrix H .

Haar-like Features - An image representation called integral image, proposed by Viola and Jones [VJ02], is used in order to calculate Haar-like features. Haar-like features calculate sums of intensities within different rectangular image regions and compute the differences between the intensity sums. Two-rectangle, three-rectangle and four-rectangle features can be applied depending on the number and position of rectangles considered.

The integral image at a point (x, y) which contributes to the sum of inten-

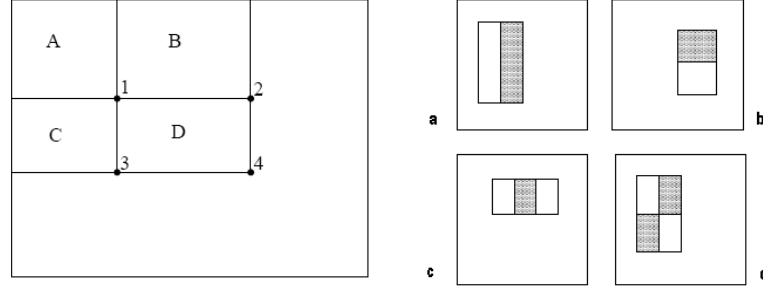


Figure 3.9: The value of integral image at location 1 is the sum of intensities in region A . The value at location 2 is $A+B$, at location 3 $A+C$ and at location 4 is $A+B+C+D$. The sum within D is equal with $4 + 1 - (2 + 3)$. (on the left); two- (a and b), three- (c) and four- (d) rectangle feature examples where the difference between gray and white rectangles is considered (on the right) [VJ02].

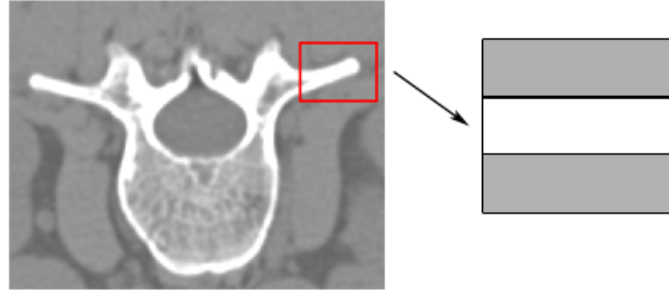


Figure 3.10: Modelling vertebra regions with Haar-like features (axial view). A transversal process can be represented by dark-bright-dark Haar-like features with three-cuboid integral volumes.

sities above and to the left of the point can be formulated as

$$II(x, y) = \sum_{\tilde{x} \leq x, \tilde{y} \leq y} I(\tilde{x}, \tilde{y}) \quad (3.18)$$

where $II(x, y)$ denotes the integral image and $I(x, y)$ is the original image. Haar-like features can be computed efficiently by integral images as any rectangular intensity sum can be calculated with 4 array references (see figure 3.9). Equation (3.18) can be applied to the three-dimensional case by additionally summing along the z -dimension of a volume. The three-dimensional

extension of the integral image can be named as integral volume. Through defining different types of Haar-like features, specific features, such as edges or lines, can be recognized. For example, transversal processes of a vertebra can be detected by Haar-like features with an integral volume combination dark-bright-dark visible on figure 3.10.

3.3.2 High-Level Feature Extraction

After extraction of low-level features at each point of a volume, high-level features are constructed based on the low-level ones. High-level features use shape and spatial information specific to a target object and thus reduce the search space by selecting only points with given high-level features and that way accelerates the solving of a MAP-MRF labelling problem on a target volume. There have been many feature detectors presented in the recent years. Among them are the Harris point, Harris-Laplace region, Harris-Affine region, Hessian-Affine region and Hessian-Laplace region detectors, which detect edges and corner like structures [MS05]. However, most of these region detectors are mainly applied on two dimensional data and their extension is difficult and/or computationally expensive to three dimensional medical data.

An example is the Difference-of-Gaussian detector for the SIFT descriptor [Low99], [AKB⁺08] which detects local scale-space extremas of the image and is a similar approach to the Hessian-Laplace region detector. To mention an example, the three dimensional version of SIFT was already applied on medical CT datasets by Allaire et al. [AKB⁺08]. It states that in case of a $512 \times 512 \times 152$ voxels CT data the computation time is 182.3 seconds for finding 160 feature points, so it is not well suitable for tasks which need a fast structure localisation of volumetric CT data [AKB⁺08].

The feature detection algorithm considered in this thesis is closely related to decision tree based supervised learning algorithms, where a function y is modelled by presenting training samples z with given features. The y values are often from a discrete set $\{1, \dots, K\}$ as these type of learning algorithms are mostly used for classification and the training samples are feature vectors with low-level features such as gradient magnitude and principal curvatures

extracted at predefined points of the volume. Based on the training examples, the learning algorithm outputs classifiers which can be used to predict the corresponding y in the target volume (e.g. if the gradient magnitude of a training example is within a certain interval then it belongs to class 1) [Die00].

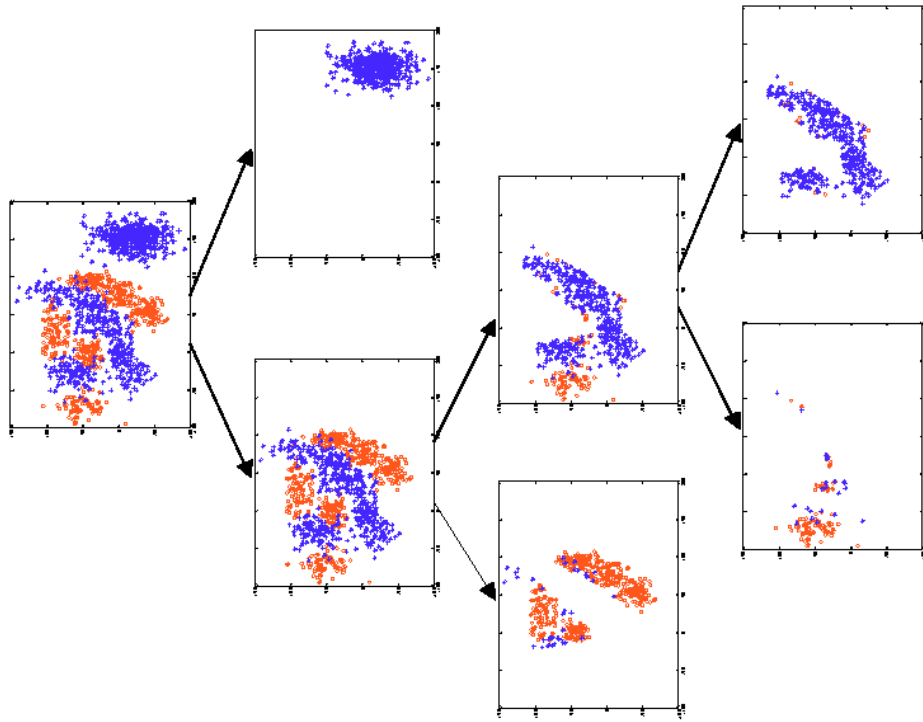


Figure 3.11: „Illustration of PBT on a synthetic dataset of 2000 points. Weak classifiers are likelihood classifiers on features such as position and distance to 2D lines. The first level of the tree divides the whole set into two parts. The right side mostly has blue (dark) points since they are away from the rest of the clouds. The tree expands on the parts where positive and negative samples are tangled.“[Tu05]

Probabilistic boosting tree (PBT) presented by Tu [Tu05] is an efficient supervised learning technique which sets up a discriminative decision process and consists of two steps for interest point detection. These two steps are learning and testing. The learning step focuses on learning a pattern of interest represented by a discriminative model $p(y|z)$, where y denotes classifications given samples z . In order to learn $p(y|z)$ the PBT framework has to be trained by presenting positive and negative samples z to the frame-

work, which is done by assigning a class y to each sample z (e.g. in the case of a binary classification a positive sample denoting foreground belongs to the class $+1$ and a negative sample denoting background corresponds to the class -1). The collection of positive and negative samples are extracted from training volumes and approximate the statistics of a certain pattern. Based on the training samples, PBT creates strong classifiers $C(z) = \sum_{j=1}^J \alpha_j c_j(z)$, which combine weak classifiers $c_j(z)$ with weights α_j on each node of the tree. Weak classifiers deliver decisions based on single features such as gradient thresholds, curvatures and Haar-like features (explained in the section 3.3.1), whereas strong classifiers merge the weak ones in a way that patterns of interest can be approximated by $p(y|z)$ as good as possible. The learned classifier on a node of the tree divides the tree into right and left nodes, which are then trained recursively.

By testing the PBT framework on a target volume, in the learning step approximated version $\tilde{p}(y|z)$ of $p(y|z)$ is used in order to classify samples extracted from a target volume. This is done by calculating conditional probabilities on each node of the tree based on the learned classifier. Left or right sub-trees are visited depending on the result of comparing the probabilities $P(+1|z)$ and $P(-1|z)$ with the two possible classes $+1$ and -1 , where the classes can represent either the underlying feature points are of interest or not (see figure 3.11).

3.4 Local Descriptors

Local descriptors describe local image regions by capturing the appearance of these regions and are extracted in both stages, MRF based model building and matching. Thus, local descriptors establish comparable measures which are necessary in order to compare local image regions around feature points delivered by a high-level feature extractor (see section 3.3.2) to those of the model objects. Furthermore, local descriptors can also describe adjacency regions between manually placed landmarks of training volumes and adjacent regions defined by feature points on target volumes respectively.

Many types of descriptors are already present in the field of machine vision, such as distribution based descriptors, which capture characteristics of

images in the form of histograms like shape context [BMP02], spin-images [JH99] and the SIFT descriptor [Low99]. Other techniques represent frequency content of images like Gabor filters or compute derivatives up to a given order like local jets [MS05]. Most of the aforementioned descriptors need special support regions such as local scale-space extremas detected by Difference-of-Gaussian (DoG) filters in the case of SIFT, mentioned in section 3.3.2.

The descriptors discussed in this section are local descriptors of manually placed landmark and automatically found feature point regions which do not require special support regions, which is important for the description of arbitrary feature points delivered by the PBT based high-level feature extraction algorithm (see section 3.3.2).

3.4.1 Vector Spin-Image

Vector spin-images were presented by Xu and Dinh [XD08] as a local descriptor on vector fields like video data, weather phenomena or medical imaging. The local descriptor of this approach reflects the statistics of the neighbourhood around arbitrary central points of vector fields. Furthermore, concepts of two well-known local descriptors, spin-images and shape contexts, are incorporated in vector spin-images.

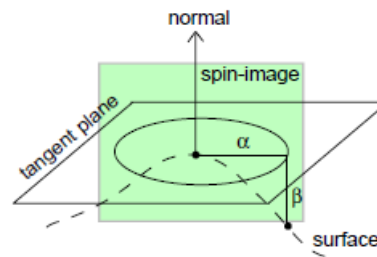


Figure 3.12: The spin-image descriptor is determined by distance α from a central point and depth β from the central point's tangent plane [XD08].

The spin-image descriptor (see figure 3.12) was originally used for three-dimensional shape based object recognition by Johnson and Hebert [JH99]. It considers surface shapes described by a dense collection of 3D points and

surface normals. In order to create a spin-image for a surface point, the orientation of the selected point has to be captured, which is done by calculating the normal of the surface at the given point by fitting a tangent plane to the surface. The spin-image descriptor is characterised by the parameters α , which is the perpendicular distance to the line defined by the surface normal at an oriented 3D point called vertex, and β , which is the perpendicular distance between the surface and the tangent plane. Finally, the spin-image descriptor is a two-dimensional histogram with dimensions α and β for vertices within a predefined support region, where each vertex in the support region is binned in the histogram according to the parameters α and β . The resulting histogram looks like an image where darker regions correspond to bins with many accumulated points and lighter regions to less frequently accumulated ones.

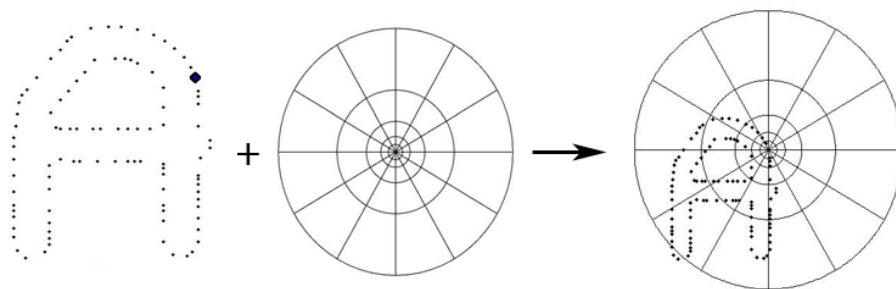


Figure 3.13: The shape context descriptor samples edge points of the object (of the letter "A" in this case) and stores the distribution of edge points related to a given point in log-polar histogram bins [BMP02].

Shape contexts (see figure 3.13) were firstly presented by Belongie et al. [BMP02] and are used for shape matching based on corresponding points on shapes which are to be matched. The distribution of the contour points of a shape are captured relative to a reference point by creating a histogram of the vectors pointing to the points of the shape relative to the reference point. Log-polar histogram bins are used in order to make the descriptor more sensitive to points which are close to the reference point than those

which are far away. Matching is finally done by finding the most similar point to the reference point on the second shape. The two points can be compared based on dissimilarity of descriptors of both points.

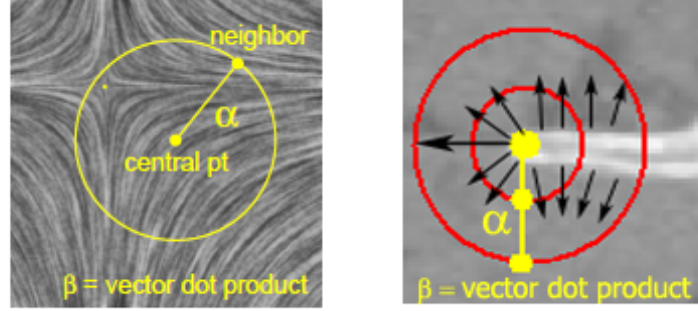


Figure 3.14: The vector spin-image descriptor [XD08] with α as the distance from the central point and β as the dot product between vectors at the central point and in the neighbourhood on the left and its application at local regions of vertebra volumes. The black arrows represent gradients in concentric distance regions (red circles) around a central point on the right figure.

Vector spin-images use concepts from both the spin-image descriptor and the shape context descriptor (see figure 3.14). It computes α for the distance from a central point and β as the dot product of the vector at the central point and that of the neighbouring point at a given distance and thus represents a rotation invariant descriptor. Although Xu and Dinh [XD08] consider the descriptor for arbitrary points in vector fields, this thesis simply uses gradient vectors g instead of vector fields as the support regions of landmarks contain high edge response. Based on the α and β values, a two-dimensional histogram is built where the bin values are weighted by the gradient magnitudes given by G and a log scale is used for the radius by

$$r_i = \exp \left\{ \ln(r_{min}) + \frac{i}{I} \left(\frac{r_{max}}{r_{min}} \right) \right\}. \quad (3.19)$$

In (3.19), r_{min} is the radius of the bin containing the central point, r_{max} is the maximum distance of the descriptor and I denotes the total number of bins where i goes from 1 to I . The histogram bins are finally normalised by the total weighted number of sample points within the bin's concentric

ring, where the gradient magnitudes G are used as a weighting factor. Normalisation is necessary in order to equalise the distribution of the histogram values. It tackles, for example, incorrect differences between points close to the boundary (where data is missing) and all other points.

3.4.2 Shape Index Histogram

Shape index is a robust local shape measure originally introduced by Koenderink and van Doorn [KvD92]. The shape index is calculated based on the principal curvatures κ_1 and κ_2 which denote the maximum and minimum curvature at a given point on the surface. The shape index is defined after Dorai and Jain [DJ95] by $\Gamma_I = \frac{1}{2} - \frac{1}{\pi} \arctan \frac{\kappa_1 + \kappa_2}{\kappa_1 - \kappa_2}$ with $\kappa_1 \geq \kappa_2$. Thus, it represents a single shape indicator where every shape is described at each point on it locally by a distinct value Γ_I except for the planar shape, where $\kappa_1 = \kappa_2 = 0$ (see table 3.1).

κ_1	κ_2		
	-	0	+
-	<i>Peak</i>	<i>Ridge</i>	<i>Saddle</i>
0	<i>Ridge</i>	<i>Flat</i>	<i>Valley</i>
+	<i>Saddle</i>	<i>Valley</i>	<i>Pit</i>

Table 3.1: The principal curvatures describe shapes depending on the interaction of their values.

The shape index is scale invariant and also invariant to geometric transformations due to the local description characteristics of the principal curvatures [Nas97]. The shape index values which occur in the interval $[0, 1]$ can be sampled within predefined regions and can be accumulated into a one-dimensional histogram where the bins represent the interval of possible shape index values (see figure 3.15). Furthermore, shape index values can be accumulated into histograms within various distance regions around a point building a two-dimensional histogram as a result, which yields a similar descriptor to the shape context descriptor [BMP02].

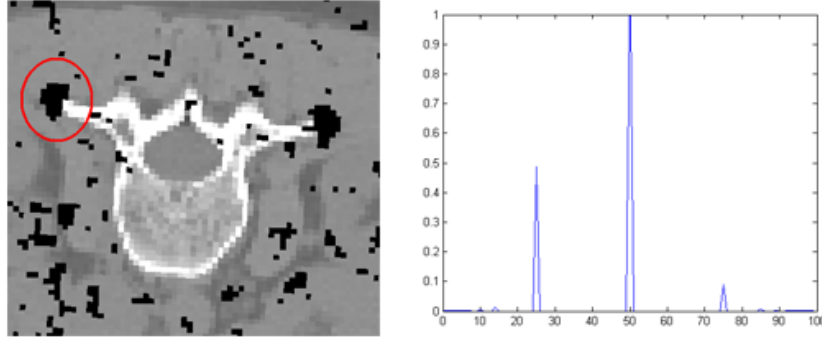


Figure 3.15: A shape index histogram of the right transversal process of a vertebra. The x-axis of the histogram represents the shape index scale and the y-axis is the proportion of points which have the same shape index. The peak in the middle corresponds to flat surfaces and the additional two peaks represent the typical shape of a transverse process.

3.5 Dissimilarity Measures

Dissimilarity Measures are essential at the MRF based model matching for preparing the probabilities in the form of qualities for solving the MAP-MRF labelling problem. Qualities are generated by comparing local descriptors of the model to those of the target volume by dissimilarity measures. In this section dissimilarity measures for the evaluated local descriptors are introduced.

3.5.1 Euclidean Distance

It is one of the most natural distance measures between two points in the n -dimensional Euclidean space \mathbb{R}^n . Here the distance is measured between two points A and B with the formula

$$d(A, B) = \sqrt{\sum_{i=1}^n (A_i - B_i)^2}. \quad (3.20)$$

By measuring histogram differences, the Euclidean distance can be considered as a bin-to-bin distance where only the difference between corresponding bins are measured in both histograms.

3.5.2 χ^2 -Divergence

Another example of a bin-to-bin distance is the χ^2 -divergence, which is a statistically relevant distance measure and is based on the χ^2 -test, where observed distributions are compared to theoretical ones. This type of distance was proposed by Xu and Dinh [XD08] as a measure for dissimilarity between vector spin-image histograms. The distance between two histograms H_1 and H_2 based on χ^2 -divergence can be calculated by

$$\chi_D^2(H_1, H_2) = \frac{1}{2} \sum_{i=1}^N \frac{(H_1[i] - H_2[i])^2}{H_1[i] + H_2[i]}. \quad (3.21)$$

3.5.3 Earth Mover's Distance

The Earth Mover's Distance is defined by Rubner et al. [RTG00] as a dissimilarity distance between signatures, where signatures are $S = \{s_j = (w_j, m_j)\}_{j=1}^N$ of size N with m_j positions and w_j weights. Histograms can be defined as special versions of signatures with this definition, where histogram values are the weights w_j and bin indices are the positions m_j of the signature. The Earth Mover's Distance (EMD) is a member of cross-bin distances, which means that no assumption about alignment of the domains of the histograms has to be made. Thus, noise and distortions inherent in histograms are considered.

EMD on two signatures $H_1 = \{(p_i, u_i)\}_{i=1}^m$ and $H_2 = \{(o_j, v_j)\}_{j=1}^n$ is defined as a transportation problem where H_1 provides the supplies at the locations u_i and H_2 the demands at locations v_j . EMD represents the minimum work for adapting the amount of supplies defined by p_i to the amount of demands defined by o_j and is calculated by the formula

$$EMD(H_1, H_2) = \min_{K=\{k_{ij}\}} \frac{\sum_{i,j} k_{ij} d_{ij}}{\sum_{i,j} k_{ij}} \quad (3.22)$$

with $\sum_j k_{ij} \leq p_i$, $\sum_i k_{ij} \leq o_j$, $\sum_{i,j} k_{ij} = \min \left\{ \sum_i p_i, \sum_j o_j \right\}$ and $k_{ij} \geq 0$ where k_{ij} are called flows which represent the amount of the i -th supply transported to the j -th demand.

3.5.4 Diffusion Distance

A recently presented alternative histogram-based distance measure to the EMD described in the section 3.5.3 is called diffusion distance [LO06]. It is one of the cross-bin distance measures for histogram-based descriptors where the difference between two histograms is considered as a temperature field, and the distance calculation happens based on the diffusion process. The integration of a norm is used as a dissimilarity measure on the diffusion field where the continuous diffusion process is discretised by Gaussian pyramids. The diffusion distance can be understood consequently as the sum of norms over all layers of the Gaussian pyramid. The diffusion distance between two histograms H_1 and H_2 is defined as

$$\hat{K}(H_1, H_2) = \int_0^{\bar{f}} k(|F(x, f)|) \partial f \quad (3.23)$$

where \bar{f} is a positive constant upper bound of the integral, and $k(\cdot)$ is a norm that measures the difference of the temperature field $F(x, f)$ to zero. An isolated temperature field is assumed where $F(x, f)$ is convolved with a Gaussian filter $\varphi(x, f)$ by $F(x, f) = F_0(x) * \varphi(x, f)$ where the natural difference between the two histograms $F_0(x) = H_1(x) - H_2(x)$ is the initial condition. The mean of the difference field converges to zero, so $F(x, f)$ becomes zero everywhere by increasing f , which can be interpreted as an equalization of H_1 and H_2 over time. Measuring the process represents dissimilarity. The Diffusion Distance is stated as a successful application to shape context, spin-image and multi-dimensional histogram based local descriptors with a similar accuracy as EMD but a much greater efficiency [LO06].

3.6 Geometrical Configuration

The geometrical configuration is setup at the MRF based model building stage and defines an adjacency system between landmarks of training volumes by Delaunay tetrahedralisation. Distances and gradient-related angles are additionally extracted along adjacencies, which contribute finally to model compatibility information. Thus, the geometrical configuration in-

incorporates the Markov-property (see section 3.2.1) into the model and describes the shape of the anatomy of interest.

3.6.1 Delaunay Tetrahedralisation

Delaunay tetrahedralisation is used to establish a neighbourhood system between training volume inherent landmarks. In order to understand Delaunay tetrahedralisation, Voronoi diagrams have to be defined. Delaunay tetrahedralisation is the three-dimensional extension of Delaunay triangulation.

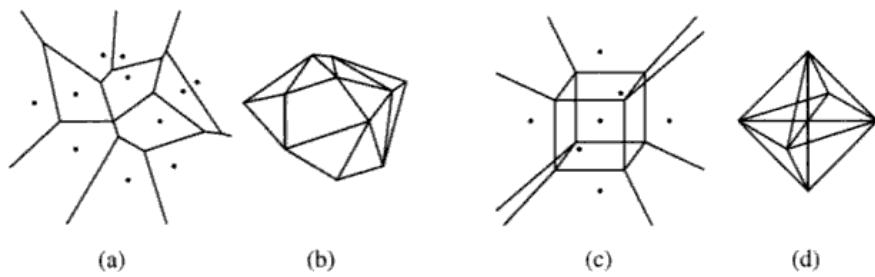


Figure 3.16: Voronoi diagram and Delaunay triangulation in two dimensions (a,b) and in three dimensions (c,d) [GO04].

A Voronoi diagram of a finite set of points in R^d called sites S partitions the space into regions, one per site $s \in S$. A region called Voronoi face for the site s contains points which are closer to s than to any other site in S . The Delaunay triangulation, the dual of the Voronoi diagram, denotes the unique triangulation where the circumsphere of every triangle in two dimensions and tetrahedron in three dimensions (where the triangles respectively tetrahedrons are built by connecting three respectively four sites) contains no sites s in its interior [GO04].

3.6.2 Distances of Adjacencies

Euclidean distances are extracted between adjacent landmarks on each training volume where landmark adjacencies are defined by Delaunay tetrahedralisation (see section 3.6.1). Lastly, the mean of the distances collected

from corresponding adjacencies of training volumes adds compatibility information to the MRF based model.

3.6.3 Gradient-Related Angles

Angles between gradient vectors g (described in section 3.3.1) and direction vectors defined by adjacent landmarks of training volumes build an additional model feature to distances mentioned in the previous section. Two angles are recorded at each compatibility based connectivity between adjacent landmarks, with each of them corresponding to an angle between the direction vector of the landmark connection defined by a compatibility $e \in E$ and the gradient vector at both landmarks of the connection. The angles are calculated as the arcus cosinus of the dot product between the normalised gradient vectors and landmark connection direction vectors by the formula

$$\beta_e = \arccos \left(\frac{\mathbf{v}_g \cdot \mathbf{v}_e}{|\mathbf{v}_g| |\mathbf{v}_e|} \right) \quad (3.24)$$

where \mathbf{v}_g is the normalized gradient vector and \mathbf{v}_e is the normalized direction vector of a landmark connection defined by a compatibility $e \in E$. Gradient-related angles enable an object-oriented description of the model and so keeps it rotation invariant as the angle representations are calculated in the range $0 - 180^\circ$.

Chapter 4

Markov Random Field Based Sparse Structure Localisation of Vertebrae

4.1 Overview

Markov Random Field (MRF) based sparse structure localisation of vertebrae is separated into MRF based model building and model matching (see section 3.1). The MRF based model building builds a model from the local appearance of vertebra landmarks (vertebra body and transverse process landmarks) and from their geometrical configuration (distances and angles along neighbourhoods). The model is built based on vertebra training data and is called sparse MRF vertebra model in this thesis as it describes the shape and texture appearance of vertebrae in a sparse manner (step 1 in figure 4.1). In the model matching part the sparse MRF vertebra model is matched with feature points detected by a high-level feature extraction algorithm on a target vertebra volume (step 2 in figure 4.1). An optimal match is delivered by a MRF solver algorithm (step 3 in figure 4.1). The matched positions provide a possible initialisation for a sophisticated model based segmentation method such as AAMs (step 4 in figure 4.1).

This chapter describes the MRF based model building and matching for the MRF based sparse structure localisation of vertebrae. For both steps the configuration of selected methods (see chapter 3) is explained.

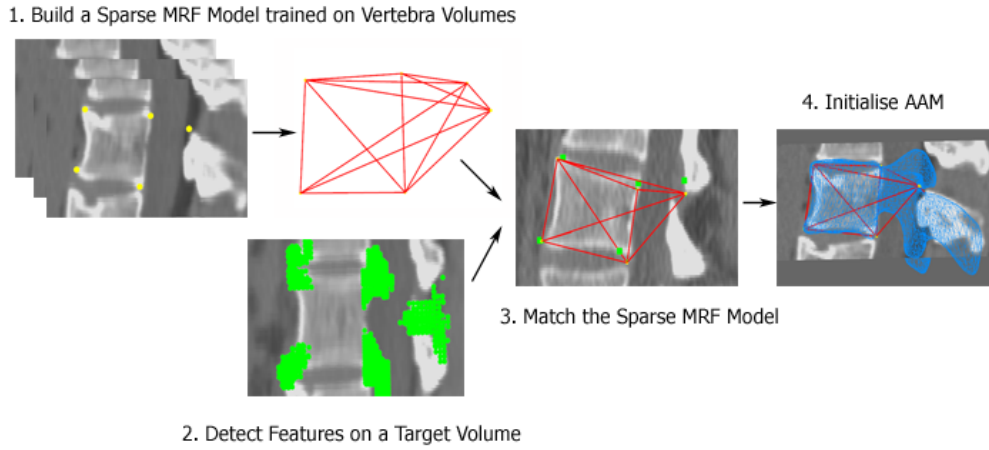


Figure 4.1: Overview of the steps of MRF based vertebra structure localisation.

4.2 Sparse MRF Model Building on Vertebrae

Building of a sparse MRF vertebra model is the first step of the MRF framework presented in this thesis. The selected methods for the sparse MRF vertebra model building and their configuration yield a rotation invariant model which makes the matching independent from the orientation of CT scans and capable of recognising disease-related rotation of vertebrae (scoliosis). This section presents, in addition to the application and configuration of the methods from Chapter 3, CT training data of vertebrae used for building sparse MRF vertebra models.

4.2.1 Training Data

In order to capture the variability of the appearance of vertebrae, 12 training volumes from 3 different datasets have been used, where each of them contains a single vertebrae and parts of the vertebra above and below with a resolution of $512 \times 512 \times 60$ and with voxel sizes of $0.25 \times 0.25 \times 1$ millimetres (mm) and $0.5 \times 0.5 \times 1$ mm (see figure 4.4). Due to the high anatomical and pathological variability of single vertebrae described in the sections 2.3 and 2.4, the restriction to a certain part of the vertebral column is necessary in order to maintain the descriptiveness and limit the variations of the sparse

MRF vertebra model. That means that every part of the vertebral column (cervical, thoracic and lumbar part) needs to have its own sparse MRF vertebra model. This thesis focuses on lumbar vertebrae, but the MRF based vertebra structure localisation framework can be extended to vertebrae of other parts.

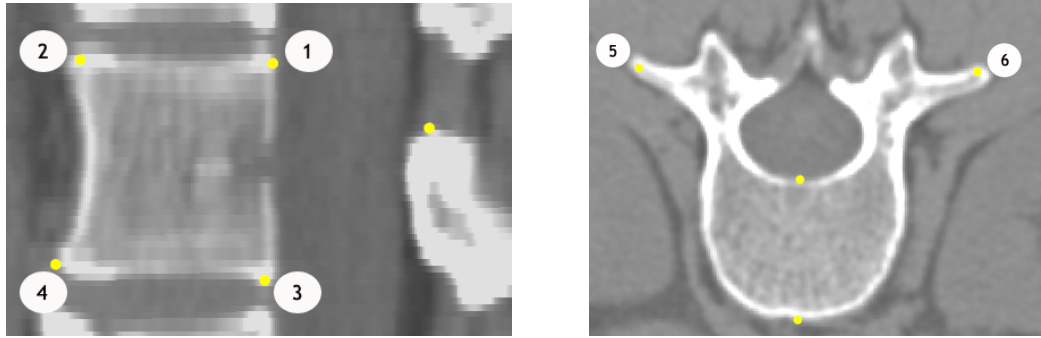


Figure 4.2: Manually placed landmarks on vertebrae from a lateral view on the left and from an axial view on the right. Landmarks 1-4 delineate the vertebra body and landmarks 5-6 represent the transverse processes.

Each of the training volumes is annotated manually at corresponding anatomically significant locations called landmarks in order to capture the sparse structure of it. It is important that each training example contains the same set of corresponding landmarks in order to achieve consistency and valuable description of the underlying anatomy. Within this thesis a configuration of 6 landmarks was selected for a sparse description of vertebrae as it was considered sufficient for model initialisation, where 4 landmarks localise the vertebra body, and additional 2 landmarks localise the transverse processes.

Figure 4.2 shows the 6 landmark positions, which are placed on the top and bottom of the anterior (point 2 and 4 in figure 4.2) and the posterior (point 1 and 3 in figure 4.2) sides and on the tips of the left and right transverse processes (point 5 and 6 in figure 4.2). This set of landmarks can describe not only the raw appearance of vertebrae, but can also be used as diagnosis related points as stated in [KHB06], where a similar but denser set of landmarks is positioned on vertebrae for feature extraction purposes (see figure 4.3). The features used in [KHB06] can denote spinal conditions, pos-

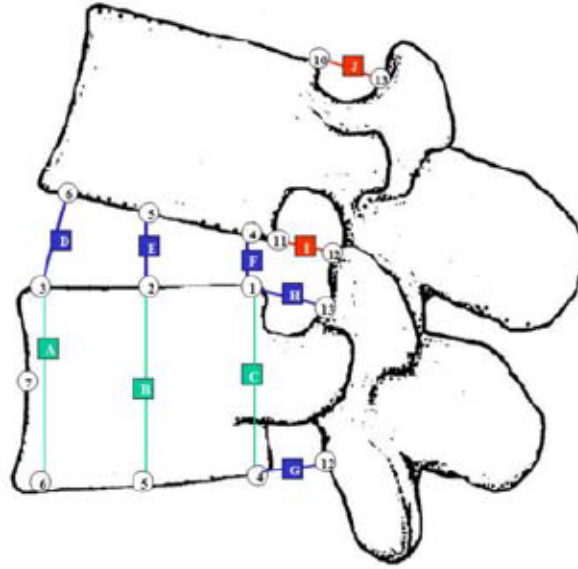


Figure 4.3: Diagnosis-relevant placement of landmarks on vertebrae. Landmarks are numbers enclosed in circles, and features extracted from the landmark configuration are represented by lines with capital letters [KHB06].

terior osteophytes, posterior apophyseal arthropathy, disc space narrowing and spondylolisthesis in order to diagnose lumbar stenosis (see figure 4.3).

4.2.2 Geometrical Configuration of the Sparse MRF Vertebra Model

One of the main aspects of MRFs is the Markov-property, which defines relations between adjacent landmarks. Thus, in order to integrate the Markov-property into the model, Delaunay tetrahedralisation provides adjacency information to the model. In addition to that, distances and gradient-related angles along adjacencies integrate additional geometrical information into the model (see section 3.6).

In the first step of the geometrical configuration construction, landmark adjacencies are established by Delaunay tetrahedralisation. For doing so, one of the training volumes is selected to define adjacencies by a Delaunay tetrahedralisation. Adjacencies derived from the selected training volume (see figure 4.5) are then established on each of the remaining 11 training volumes (see figure 4.6).

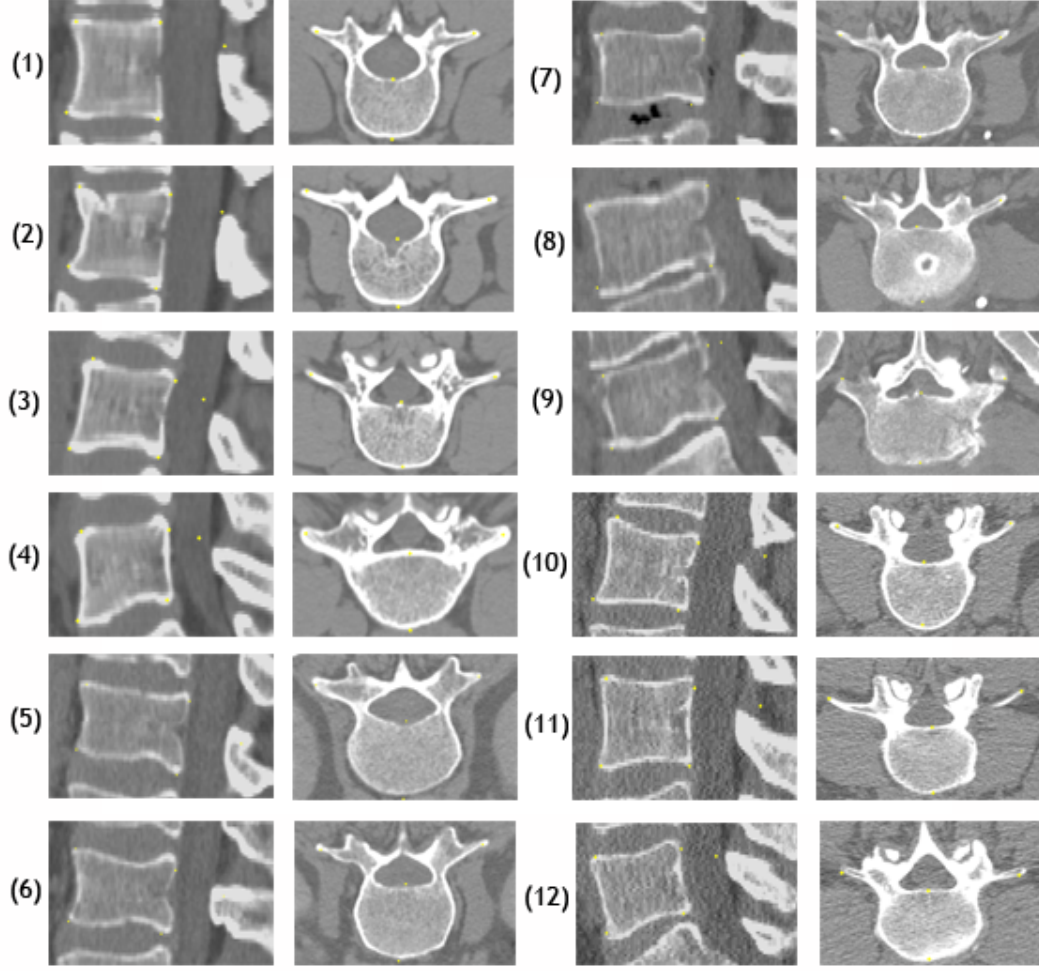


Figure 4.4: CT volumetric training data of 12 lumbar vertebrae from lateral and axial views.

After extracting adjacencies, distances l_e^i with $i = 1..12$ are calculated between corresponding landmarks on each of the i training volumes as Euclidean distances. Mean distances \bar{l}_e and their standard deviations l_e^σ are calculated from corresponding distances l_e^i over each of the i training volumes and are stored in the compatibilities $e \in E$ of the sparse MRF vertebra model. Furthermore, each landmark adjacency gets two angles assigned, β_{e1}^i and β_{e2}^i , which measure the angle between the vector represented by adjacent landmarks and the gradient vector g at landmarks of the training data

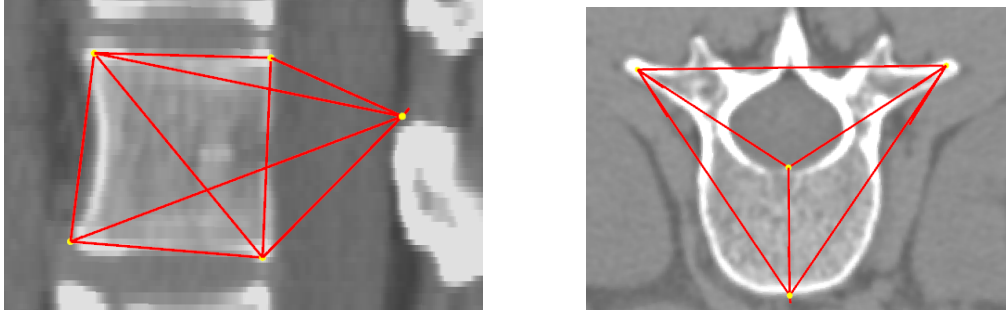


Figure 4.5: Geometrical setup of the sparse MRF vertebra model with Delaunay tetrahedralisation from the lateral view on the left and from the axial view on the right.

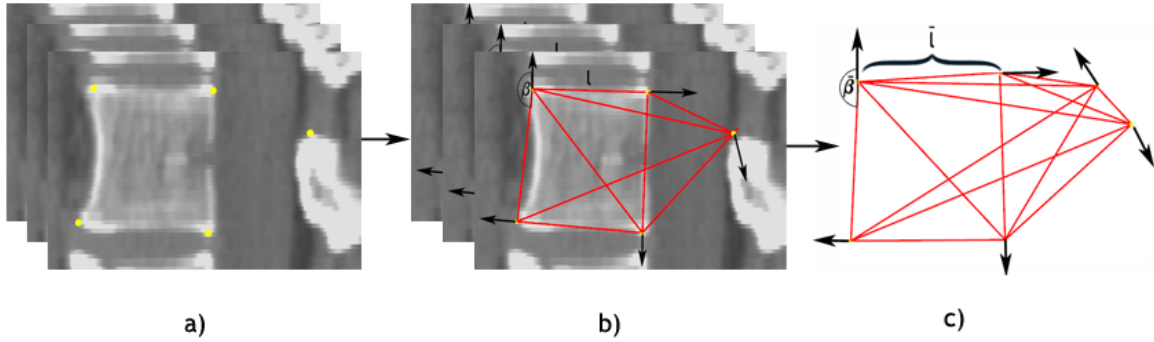


Figure 4.6: a) Adjacency information is extracted from manually annotated (landmarks as yellow points) training volumes and builds the model compatibilities. b) Distances l between adjacent landmarks and gradient-related angles β are extracted from each training volume based on the Delaunay tetrahedralisation (red lines). c) Mean distances \bar{l} and mean angles $\bar{\beta}$ are computed from the training volumes for each compatibility $e \in E$. The black arrows at each landmark represent gradient vectors.

(see section 3.6.3). The two angles are calculated over corresponding landmark adjacencies, and their mean angles $\bar{\beta}_{e1}$ and $\bar{\beta}_{e2}$ and standard deviations β_{e1}^σ , β_{e2}^σ are stored in the compatibilities $e \in E$ of the model in addition to the distances (see figure 4.6). Through these parameters the model captures the underlying structure and represents it in a rotation invariant manner.

4.2.3 Local Descriptor Configuration of the Sparse MRF Vertebra Model

Local appearance of regions around vertebra landmarks is represented by local descriptors. Local descriptors $D_t^i, i = 1..12$ are built in regions around corresponding landmarks in the i training volumes. For the sparse MRF vertebra models, histogram-based local descriptors are used, which are useful techniques to capture the distribution of appearance- and shape-related features at local image regions. The histogram-based local descriptors described in section 3.4 are applied to MRF based sparse vertebra structure localisation.

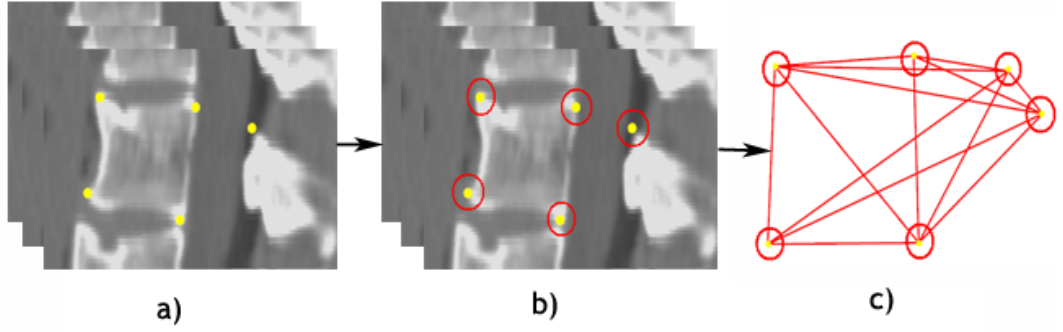


Figure 4.7: a) Local support regions are extracted around landmarks (yellow points) of the training volumes. b) Local descriptors are constructed within local support regions with predefined spherical distances. c) Means of the local descriptors \bar{D}_t are computed considering corresponding landmarks on all training volumes which finally deliver model object $t \in T$ information.

Local descriptors $D_t^i, i = 1..12$ are constructed around corresponding landmarks over all of the i training volumes. Mean \bar{D}_t descriptors are recorded from local descriptors $D_t^i, i = 1..12$ of corresponding landmarks over the i training volumes [DMLB07] (see figure 4.7). The mean descriptors \bar{D}_t describe finally the objects $t \in T$ of the sparse MRF vertebra model. This means that, for example, local descriptors are constructed around transverse process landmarks corresponding to object $t_1 \in T$ on each training volume, and the mean of them is finally computed and accounts for the description of the object t_1 .

The local support regions around landmarks are represented by spherical regions with a diameter of 12 mm. The local descriptors are constructed within these regions and represent vertebra appearance in a rotation invariant manner. Three types of local descriptors (see section 3.4), vector spin-images, shape index histograms and gradient-based edge descriptors are considered for MRF based sparse structure localisation of vertebrae in this thesis where vector spin-images and shape index histograms describe local landmark regions and gradient-based edge descriptors extract additional information along landmark connections.

Vector Spin-Image - Vector spin-images were tested with similar parameters as stated in the paper [XD08]. The final configuration of $r_{min} = 2$ mm and $r_{max} = 6$ mm with 10×10 two dimensional histograms was used.

Shape Index Histogram - Shape index histograms were represented by a one-dimensional histogram with 100 bins, where each bin corresponds to a quantised shape index from the range $[0, 1]$ to $[0, 100]$, and their values contain the number of shape indices of a certain type in the support region. Two-dimensional versions of shape index histograms were also tested, where shape index values are gathered within different spherical distance regions. For distance calculation the formula (3.19) was used.

Gradient-based Edge Descriptor - It was experimented with extraction of image gradient information along landmark connections according to the compatibilities of the model. Gradients g were sampled at 20 positions along landmark connections as most of the landmark connections run along borders of the underlying vertebrae. The gradient magnitudes G and orientations relative to vectors defined by landmark connections were recorded at the sampled positions. The mean values of the gradient information are calculated at each of the 20 positions considering all training volumes and are stored in each compatibility $e \in E$.

4.3 Sparse MRF Vertebra Model Matching with Target Data

4.3.1 Overview

Given the sparse MRF vertebra model and a target vertebra volume, the best fit match of the sparse model with the feature points detected by a high-level feature extraction algorithm is of interest where the feature points are the possible labels for the model objects. The best fit match, which localises a target vertebra sparsely, can be found by an MRF solver algorithm. It represents a feature point configuration with maximal qualities based on the comparison of local descriptors and geometrical configuration of the sparse MRF vertebra model to those of feature points on the target data (see figure 4.8).

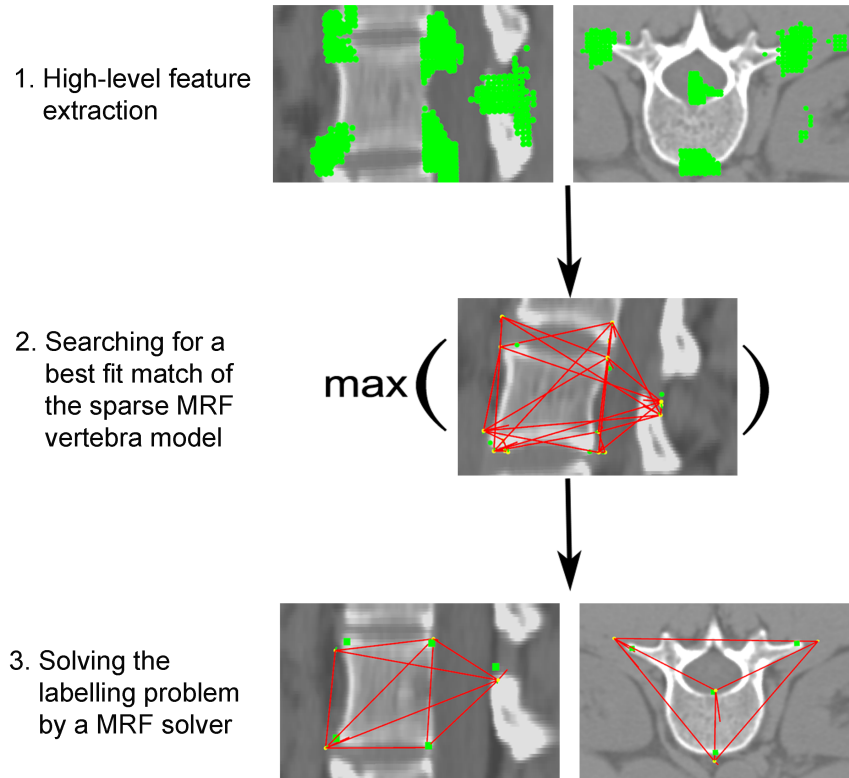


Figure 4.8: MRF model matching on target vertebra volumes consists of the steps 1-3 visible in the figure.

4.3.2 High-Level Feature Extraction in the Vertebra Target Data

The MRF vertebra model is matched on feature points detected by the probabilistic boosting tree (PBT) approach described in section 3.3.2, which is trained with positive and negative examples of landmarks of the training volumes described in section 4.2.1. Applying the algorithm on a target volume, feature points can be detected with the help of the trained classifiers.

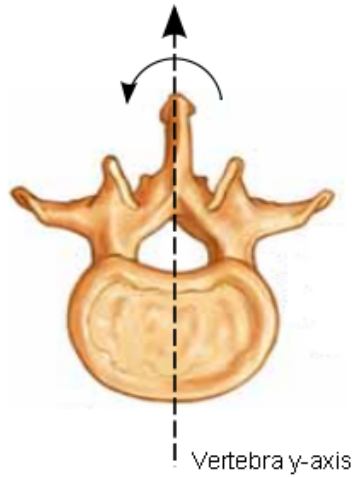


Figure 4.9: The high-level feature extraction algorithm delivers feature points for the vertebra body anterior top and bottom part jointly which is also the case with the top and bottom part of the posterior side. Thus, a rotation invariant matching of the vertebra body part of the sparse MRF vertebra model around the y-axis of the vertebra is possible. Additional detection of the left and right transverse processes jointly enables rotation invariant matching of the whole sparse MRF vertebra model. [Uni10]

The high-level feature extractor delivers k_t feature point candidates called labels \hat{x} for each of the $|T| = 6$ landmarks modelled by the objects of the sparse MRF vertebra model which yield $k_{t_1} * k_{t_2}$ edges between feature points that are possible labels for landmarks modelled by the objects t_1 and t_2 . Thus, each of the landmarks 1-6 in figure 4.2 get clouds of feature points assigned. However, it has to be remarked that the landmarks 1 and 3 and respectively 2 and 4 on figure 4.2 get the same cloud of feature points assigned. This is due to the fact that the feature extractor jointly recognizes the vertebra

body anterior top and bottom part (landmarks 2 and 4 on figure 4.2) which is also the case with the vertebra body posterior top and bottom part. This behaviour allows rotation invariant matching of the vertebra body around the object oriented y-axis of the vertebra (see figure 4.9). In cases where the feature extractor delivers feature points for both the left and right transverse processes (landmarks 5 and 6 on figure 4.2) jointly, the whole sparse MRF vertebra model matching becomes rotation invariant around the object oriented y-axis of the vertebra. Values for k_t are assumed to be up to 600 for performance reasons within this thesis.

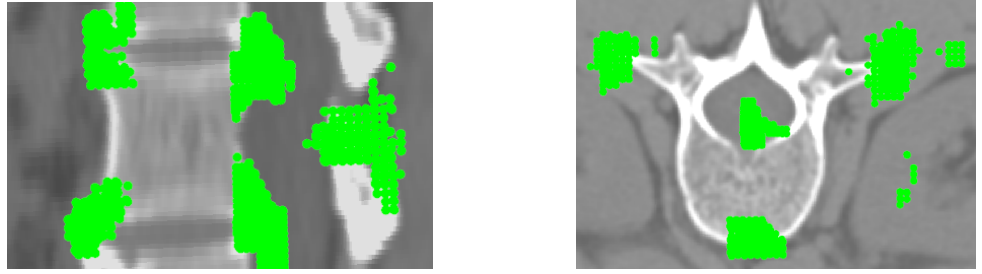


Figure 4.10: Results of the PBT based feature extraction algorithm for the modelled landmarks of a vertebra where the green dots denote feature points.

As a next step, local descriptors $D_{\hat{x}}$ are extracted around each label \hat{x} , where each label is a possible candidate for the object $t \in T$ which models the landmark for which the PBT based high-level feature extractor delivered that particular label. That way, each label descriptor is compatible with a certain object descriptor \bar{D}_t .

Furthermore, distances l_{a_e} and angles $\beta_{a_{e1}}, \beta_{a_{e2}}$ are extracted along edges $a_e \in A_e$ with $a_e = \{(t, \hat{x}), (t', \hat{x}')\}$ between adjacent labels where adjacency is defined by the compatibilities $e \in E$ of the sparse MRF vertebra model.

4.3.3 Quality Assignment for the Sparse MRF Model Matching

Having available the local descriptor and geometrical configuration information from the sparse MRF vertebra model as well as from labels detected by the high-level feature extraction on target data, qualities for model-to-target matches can be computed. Qualities indicate approximations for posterior

probabilities and make assumptions about how likely it is to match the model on a certain configuration of labels (see section 3.5).

Label qualities $q_t(\hat{x})$ are calculated based on dissimilarity between the local descriptors of the model objects \bar{D}_t and the feature point descriptors $D_{\hat{x}}$ of the target volume. For quality assignment between labels and object descriptors, the shape index histogram and the vector spin-image descriptor are used in the same configuration as stated in the section 4.2.3. As dissimilarity measures, the methods in section 3.5 are considered. The local descriptors, shape index histograms and vector spin-images were generally tested with the χ^2 -divergence measure as proposed by Hetzel et al. [HLLS01] and Xu and Dinh [XD08]. Furthermore, both descriptors were also tested with the diffusion distance and EMD (see section 3.5), as these measures are robust to deformations and noise.

Finally, the qualities are obtained by normalising the results based on the dissimilarity measures by

$$q_t(\hat{x}) = -\text{norm}(D_{\hat{x}} - \bar{D}_t) \quad (4.1)$$

where $\text{norm}(\cdot)$ means the normalisation by the medians of the labels of each object $t \in T$ in order to be invariant to outliers. Thus, the qualities are normalised to a scale where values have a median of -1 and a maximum of 0 [DMLB07], [DWBL09].

The edge qualities $q_{tt'}(\hat{x}, \hat{x}')$ are calculated between the Gaussian distributions of compatibility distances \bar{l}_e, l_e^σ and angles $\bar{\beta}_{e1}, \beta_{e1}^\sigma, \bar{\beta}_{e2}$ and β_{e2}^σ of the sparse MRF vertebra model and the corresponding edge distances l_{ae} and angles β_{ae1} and β_{ae2} through the formulas

$$\begin{aligned} l_{q_{tt'}} &= \text{norm}\left(1 - \exp\left(-\left(l_{ae} - \bar{l}_e\right) / \left(2 * \bar{l}_e^{\sigma^2}\right)\right)\right) \\ \beta_{q_{tt'}}^{e1} &= \exp\left(-\left(\beta_{ae1} - \bar{\beta}_{e1}\right) / \left(2 * \bar{\beta}_{e1}^{\sigma^2}\right)\right) \\ \beta_{q_{tt'}}^{e2} &= \exp\left(-\left(\beta_{ae2} - \bar{\beta}_{e2}\right) / \left(2 * \bar{\beta}_{e2}^{\sigma^2}\right)\right) \\ \beta_{q_{tt'}} &= \text{norm}\left(1 - \left(\omega \cdot \beta_{q_{tt'}}^{e1} + v \cdot \beta_{q_{tt'}}^{e2}\right)\right) \\ q_{tt'}(\hat{x}, \hat{x}') &= -\left(\varphi \cdot l_{q_{tt'}} + \gamma \cdot \beta_{q_{tt'}}\right) \end{aligned} \quad (4.2)$$

where the final edge quality $q_{tt'}(\hat{x}, \hat{x}')$ is a weighted sum of the distance qualities and the angle qualities [DWBL09]. The setting $\varphi = \frac{1}{2}$ and $\gamma = \frac{1}{2}$ was used, thus taking the average of the two factors. The same setup is considered for ω and v . The function $norm(\cdot)$ here also means the normalisation by the median, however in this case the range $[-1, 0]$ is considered. Edges with a lower quality than that of the median are not considered for the MRF matching, which means that edges with lower values than -1 are not considered for possible candidates of compatibilities and are discarded [DMLB07].

In cases where the gradient-based edge descriptor is used, the formulas in (4.2) are extended by

$$D_{q_{tt'}} = norm(D_{a_e} - D_e) \quad (4.3)$$

where $norm(\cdot)$ is considered again as the median normalisation function, where lower edge qualities than -1 are cropped again. The Euclidean distance was used as a dissimilarity measure between D_{a_e} and D_e (see section 3.5). The overall edge quality measure is modified to

$$q_{tt'}(\hat{x}, \hat{x}') = - \left(\varphi \cdot l_{q_{tt'}} + \gamma \cdot \beta_{q_{tt'}} + \eta \cdot D_{q_{tt'}} \right) \quad (4.4)$$

in this special case where again the average is considered with $\varphi = \gamma = \eta = \frac{1}{3}$. In addition to that, more guidance by the distance and the edge descriptor turned out to be useful, for instance by considering parameter settings such as $\varphi = \eta = \frac{2}{5}$ and $\gamma = \frac{1}{5}$.

It can happen that the high-level feature extraction algorithm does not deliver feature points for the modelled landmarks where the model would expect candidates. For these special cases artificial points are added manually which describe the objects of the sparse MRF model where feature point information is missing in order to maintain the matching activity [DMLB07].

4.3.4 Solving the MAP-MRF labelling problem with the Max-sum Solver

The Max-sum approach was selected to solve the vertebra MAP-MRF labelling problem due to its successful application to sparse structure localisation tasks of the same kind [DMLB07]. The Max-sum solver presented already in the

section 3.2.2.1 is based on the augmenting DAG approach [Wer07] and solves the MAP-MRF labelling problem by searching for a label configuration which maximises the qualities of the label-to-object and edge-to-compatibility matchings. The result is a maximum quality labelling with 6 labels where each label is matched with one of the 6 model objects which describe landmarks on a target volume. At first, the augmenting DAG approach is explained, followed by its setup for solving the vertebra labelling problem.

Augmenting DAG Algorithm - The technical report of Werner [Wer05] describes the augmenting directed acyclic graph (DAG) algorithm for solving the Max-sum problem by decreasing its upper bound. A DAG is a directed graph with no cycles, where a directed graph is a graph whose edges have directions. It is acyclic, which means that if following a path by starting at a vertex v , there are no paths to loop back to v again.

The main idea of the augmenting DAG algorithm is to solve the Max-sum problem by minimising the problem height with equivalent transformations (see section 3.2.2.1). For doing so, a special arc-consistency algorithm is run on maximal nodes and edges as a first step in order to eliminate the maximal nodes and edges which are not arc-consistent (see section 3.2.2.1). Pointers to nodes which caused eliminations are stored at each eliminated node forming a DAG. If all labels of an object are eliminated, the pointers are backtracked resulting in an augmenting DAG. Along the augmenting DAG equivalent transformations, which decrease the problem height, are finally done in order to move towards the optimal solution. A short overview of the iteration steps of the augmenting DAG algorithm [Wer05] is given here:

1. The *first step* starts a special arc-consistency algorithm on maximal labels and edges as follows: an auxiliary variable $b_t(\hat{x})$ is set to each label (t, \hat{x}) as $b_t(\hat{x}) := ALIVE$ if (t, \hat{x}) is maximal and $b_t(\hat{x}) := NONMAX$ if (t, \hat{x}) is not maximal. If a pencil (t, t', \hat{x}) is found with $b_t(\hat{x}) = ALIVE$, but violating the condition

$$\exists \hat{x}' \left[\text{edge} \left\{ (t, \hat{x}), (t', \hat{x}') \right\} \text{ maximal, } b_{t'}(\hat{x}') = ALIVE \right], \quad (4.5)$$

then label (t, \hat{x}) is deleted by setting $b_t(\hat{x}) := t'$. This step is repeated

until no such pencils exist, or an object t^* is found with $b_{t^*}(\hat{x}) \neq \text{ALIVE}$ for all $\hat{x} \in \mathcal{L}$. In the former case the augmenting DAG algorithm stops; in the latter case, it is continued to the *second step*.

After every iteration of this step, the maximal edges and the variables $b_t(\hat{x})$ set up a directed acyclic sub-graph D of the graph $(T \times \mathcal{L}, A_E)$. If t^* has been found, the augmenting DAG $D(t^*)$ represents a sub-graph of D .

2. In the *second step*, the search direction $\Delta\mu$ of the height decrease with the condition $U(q^{\mu+\lambda\Delta\mu}) < U(q^\mu)$ is found, where q^μ states a max-sum problem after an equivalent transformation (see section 3.2.2.1). The smallest vector $\Delta\mu$ is found by traversing the DAG $D(t^*)$ created in the previous step.
3. In the *third step*, the length λ of the search step is found by checking the conditions that no edge becomes positive, the height of no object is increased and the height of t^* is decreased. This step is completed by the equivalent transformation $\mu = \mu + \lambda\Delta\mu$ which induces that some non-maximal nodes and edges will become maximal and some maximal edges non-maximal. Thus, the DAG is updated and the augmenting DAG algorithm is continued with the next iteration by going to the *first step*.

A threshold parameter θ is introduced to the augmenting DAG algorithm which should control the height decrease and force taking high enough λ values in order to be able to terminate in a finite number of iteration steps. Thus, selecting large θ values the maximality of labels and edges increases resulting in larger searching steps λ .

Furthermore, it has to be remarked that the augmenting DAG algorithm can not solve the Max-sum problem always to optimality (already mentioned in section 3.2.2.1), because an arc-consistent set of labels as a result of the algorithm is necessary but insufficient for a minimum problem height $U(q)$ [Wer07], [Wer05]. This means, that the algorithm can terminate with more than one label per object resulting in an ambiguous matching of the sparse MRF vertebra model.

Configuration of the Augmenting DAG Algorithm - The Max-sum augmenting DAG algorithm solves the MAP-MRF labelling problem by the function

$$\mathbf{L} = \text{maxsum}(\Omega, \rho, \Phi\Phi, \Phi, \theta) \quad (4.6)$$

with the following set of parameters:

- Ω is a matrix where each column denotes a vector of indices of neighbouring object pairs t and t' which define a compatibility $e \in E$. This information is extracted from the geometrical configuration of the sparse MRF vertebra model.
- ρ is a vector which has $|T|$ elements. An element $\rho(t)$ stores the number of labels of object t .
- $\Phi\Phi$ is a matrix which contains the edges a_e and their qualities as column vectors in the form $(e, (t, \hat{x}), (t', \hat{x}'), q_{tt'}(\hat{x}, \hat{x}'))$. The edge qualities $q_{tt'}(\hat{x}, \hat{x}')$ contain normalised values in the interval $[-1, 0]$.
- Φ is a vector with $\sum \rho$ elements which contains qualities of each label $q_t(\hat{x})$ for each object t . Thus, Φ contains a sequence of label qualities where the first k_{t_1} elements denote the qualities of the first object $t_1 \in T$ if it has k_{t_1} feature point candidates. The label qualities $q_t(\hat{x})$ are normalised to having a maximum of 0 and median of -1 .
- θ is a tolerance threshold for the relaxation labelling.

The Max-sum algorithm ultimately delivers a MAP configuration of the MRF in the form of a logical vector \mathbf{L} , which has the same setup and size as the vector Φ , where each position corresponds to those of Φ . The MAP labelling result of the algorithm can be found by traversing the logical vector \mathbf{L} consisting of ones and zeros and looking for live labels by seeking for ones. Through the information in ρ and the index of the ones, the corresponding objects t of the labelling can be found.

The parameter θ is a component which enhances the relaxation by forcing the algorithm to take high enough iteration steps towards the optimal labelling (see section 3.2.2.1). Thus, with $\theta > 0$ the algorithm terminates in a finite number of iterations. θ was set to 2^{32} in this thesis.

4.4 Implementation

The MRF based vertebra sparse structure localisation framework consists of two major parts, the sparse MRF model builder and the sparse MRF model matcher.

Sparse MRF Model Builder - It was implemented in Java. Local descriptor information is extracted from landmark regions of the training data, and distances and gradient-related angles are recorded between adjacent landmarks on each training volume. The object and compatibility information of the model represented by mean descriptors, distances and angles calculated over all training volumes is stored in XML format. The local descriptors presented in section 3.4 were implemented in Java. Adjacency is established with the Delaunay tetrahedralisation on a preselected training volume. For the Delaunay tetrahedralisation, the CGAL [CGA10] implementation was used.

Sparse MRF Model Matcher - It was implemented in Java. The sparse MRF model matcher collects object and compatibility information from the XML file of the sparse MRF model and compares them to label and edge information extracted from a target vertebra volume by a PBT based feature extraction technique (see section 3.3.2). The PBT based feature extraction algorithm was provided by VRVis [VRV10] and was implemented in Java. The comparison between local descriptors of the sparse MRF model and those of the feature points detected by the PBT based feature extraction algorithm was done by dissimilarity measures (see section 3.5). The Euclidean distance, χ^2 -divergence and Diffusion distance were implemented in Java. For the Earth Mover's Distance, a C-implementation by Y. Rubner [Rub99] was used. For solving the MAP-MRF labelling problem, the augmenting DAG based Max-sum solver from T. Werner [Wer05] was applied, which was implemented in C [Wer09].

Chapter 5

Results

5.1 Evaluation Framework

Localisation of the structure of anatomies can be considered as a raw segmentation by labelling characteristic parts of the given anatomy. This section describes the data based on which the labelling was done. Furthermore, it is explained which information was used as a reference for the evaluation of the results.

Data - The evaluation of the sparse MRF vertebra model matching was done on the 12 lumbar vertebrae training volumes described in section 4.2.1.

Ground Truth - In order to evaluate the results of the labelled data, a corresponding exact segmentation of the target anatomy is required. This manual segmentation is called the ground truth, which is prepared by delineating the anatomy through a physician. In a raw segmentation case, a manual labelling of the anatomically significant points is necessary by placing landmarks at the required locations. The landmark information is extracted from the annotated training data described in section 4.2.1. The result of the labelling is evaluated by generating the distance between the matched labels and the manually placed landmarks.

5.2 Error Measures

For the evaluation of the sparse segmentation results, the quantitative leave-one-out method was used for validating the model through the training data. With the leave-one-out strategy, a volume was chosen as a target volume, and the model was built on the remaining 11 volumes. Thus, it was tested how well the sparse MRF vertebra model can describe the data to be matched if it is left out of the model. The accuracy of the matches was measured by the distance between landmarks used for building the sparse MRF vertebra model and the labelled locations. The matching error D_{mt} was measured as the average of Euclidean distances between each point of the ground truth and of the matching result by the formula

$$D_{mt}(\Lambda_{mt}, \Lambda_{gt}) = \frac{1}{n} \sum_{i=1}^n \sqrt{(x_{mt,i} - x_{gt,i})^2 + (y_{mt,i} - y_{gt,i})^2 + (z_{mt,i} - z_{gt,i})^2} \quad (5.1)$$

where Λ_{mt} represents the matched label positions in the target volume and Λ_{gt} stays for landmarks of the ground truth of the target volume. The average is taken from the results by dividing by n which is 6 here in order to be able to calculate the error related to all 6 ground truth positions. Average matching errors D_{mt} are also calculated separately based on the 4 ground truth positions of the vertebra body and on the 2 ground truth positions of the transverse processes. This is done due to the fact that whenever the vertebra body is localised with a good accuracy, it can already be taken as an appropriate initialisation for a subsequent full segmentation step.

The results were generated with the two local descriptor approaches: one-dimensional shape index histograms and vector spin-images described in section 3.4. The geometrical configuration remained the same in both cases using Euclidean distances and measuring gradient-related angles for the model-to-target matches. Gradient-based edge descriptors (see section 4.2.3) were applied in an additional test.

5.3 Evaluation of the Results

Matching errors with leave-one-out tests on the training volumes are visible on the figures 5.3, 5.4 and 5.5. High matching errors occurred on the training volumes 3, 4, 8, 9 and 12, which represent either the 5th lumbar vertebra (L5) or the 4th lumbar vertebra (L4). Thus, the matching delivered more accurate results on upper lumbar vertebrae (L1-L3) than on vertebrae of lower regions, such as L4 and L5. The L4 and L5 vertebrae volumes contained bony structures similar to the tips of transverse processes in the vicinity of their ground truth landmarks. Thus, the algorithm considered structures such as the top of the hips as a good fit for these positions resulting in mismatches.

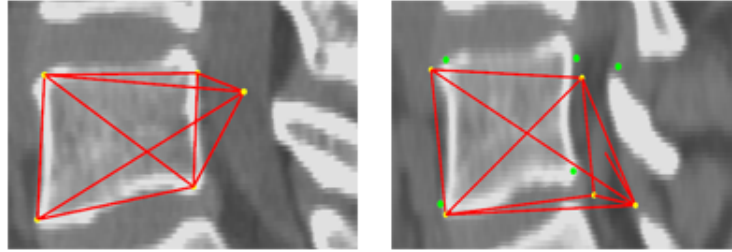


Figure 5.1: The L5 sparse MRF vertebra model (on the left) and its rotated matching (on the right). For the left and right transverse processes parts of the hip were matched causing a rotated matching of the vertebra body.

Mismatches caused also results where rotated variants of the sparse MRF vertebra model were matched to the target volume. It was possible because the model represents vertebrae in a rotation invariant manner around its y-axis (see section 4.3.2). It happened, for example, at L4 vertebrae where the transverse processes of the subsequent L5 vertebra were close enough to match the processes on the L5, that a rotation variant of either the vertebra body or the whole model was matched. In cases of rotated vertebra body matchings the labelling of the upper and lower landmarks described by model objects of the vertebra body were switched. In cases where the high-level feature extractor selected feature points for the left and right transverse processes jointly, rotated versions of the whole model could occur.

This means that in addition to the switched vertebra body upper and lower regions the left and right transverse processes were also matched the opposite way around. These phenomena are well visible in figure 5.1 and result histograms 5.3, 5.4 and 5.5.

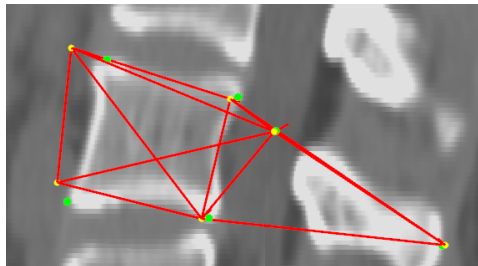


Figure 5.2: Adding the spinous process as an additional landmark to the sparse MRF vertebra model.

Furthermore, ambiguous matching results appeared in approximately 15% of all test cases, which is due to the fact that the augmenting DAG algorithm had termination states with non-minimal upper bounds (see section 4.3.4). This means that more than one feature point was matched with some of the objects of the sparse MRF vertebra model. In these cases the mean of the selected feature points per object were calculated as a result.

An additional solution for ambiguous matchings is stated by Werner [Wer07] and Donner et. al [DMLB07] where extra landmarks are added to the model in order to disambiguate the matching. So adding extra landmarks to the sparse MRF vertebra model and detecting them on target vertebra volumes have to be investigated and tested as a future work. A possible model extension is placing a landmark on the spinous process (see figure 5.2). Adding extra model landmarks could be also helpful to avoid rotated vertebra matchings at L4 and L5 vertebrae.

The MRF framework on both the one-dimensional shape index histograms and vector spin-images delivered similar results, which is apparent by comparing figures 5.3 and 5.4. Testing with one-dimensional shape index histograms resulted in a average matching error of 5.43 mm for the L1-L3 vertebrae and 10.73 mm for the L4-L5 vertebrae. The average matching errors for vector spin-image matching are 5.52 mm for the L1-L3 vertebrae and 9.51

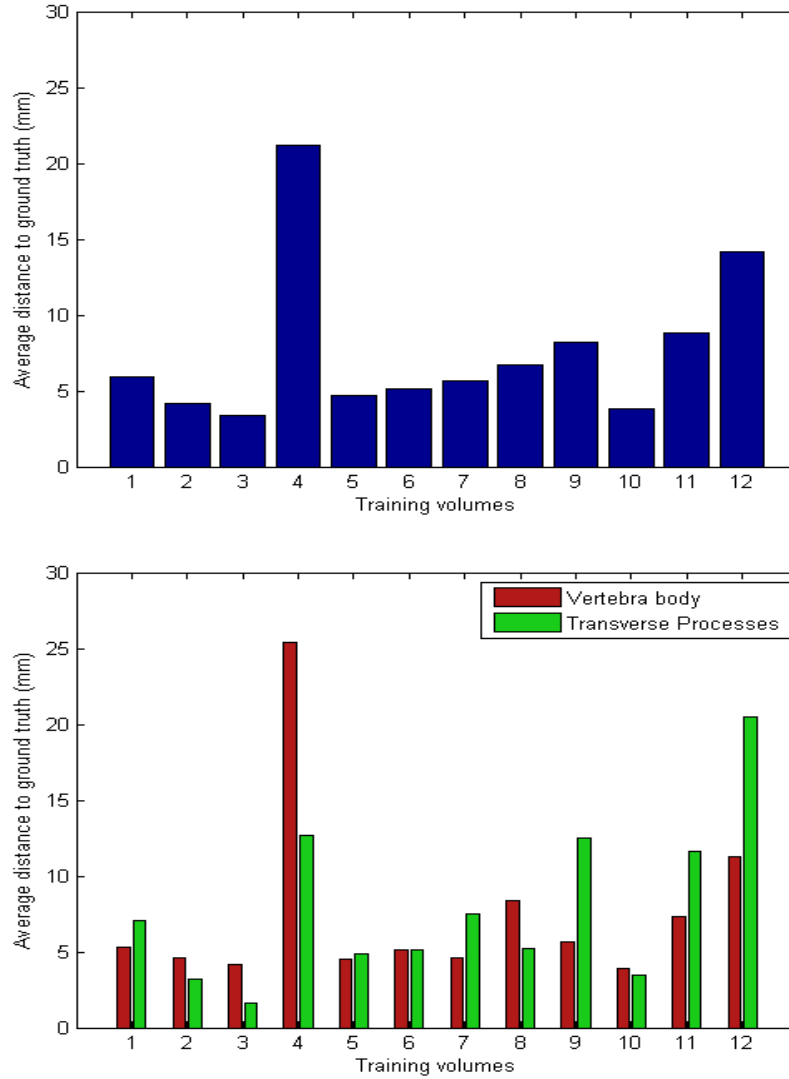


Figure 5.3: Average matching errors based on all 6 anatomical landmarks of the sparse MRF vertebra model (on the top) and average matching errors based on the 4 landmarks of the vertebra body and the 2 landmarks of the transverse processes (on the bottom) by the leave-one-out test with one-dimensional shape index histograms applied on the 12 lumbar vertebra training dataset.

mm for the L4-L5 vertebrae.

Finally, it was experimented with additional edge descriptions of the sparse MRF vertebra model in order to improve the accuracy of the matching results. In addition to the components of the geometrical configuration of the model

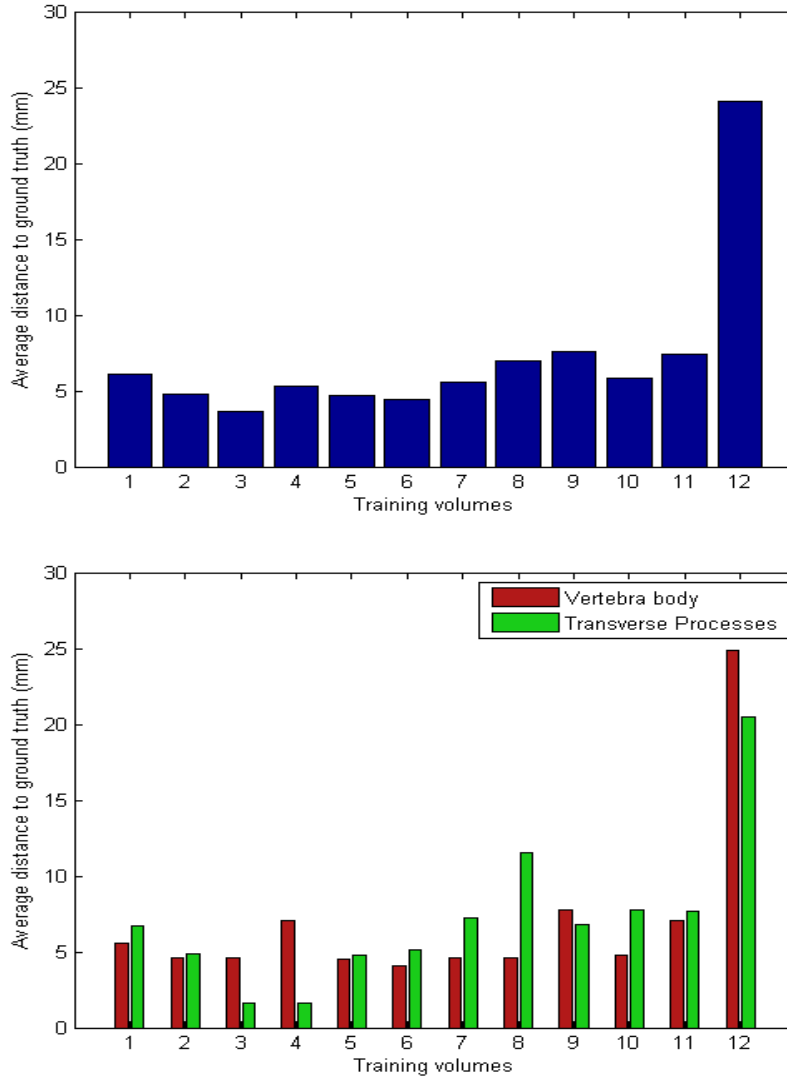


Figure 5.4: Average matching errors based on all 6 anatomical landmarks of the sparse MRF vertebra model (on the top) and average matching errors based on the 4 landmarks of the vertebra body and the 2 landmarks of the transverse processes (on the bottom) by the leave-one-out test with vector spin-images applied on the 12 lumbar vertebra training dataset.

(see section 3.6), descriptors along edges were extracted.

Thus, gradient orientations relative to the landmark adjacency vectors were sampled at 20 positions and its means were stored in the compatibilities of the sparse MRF vertebra model (see section 4.2.3). At the model

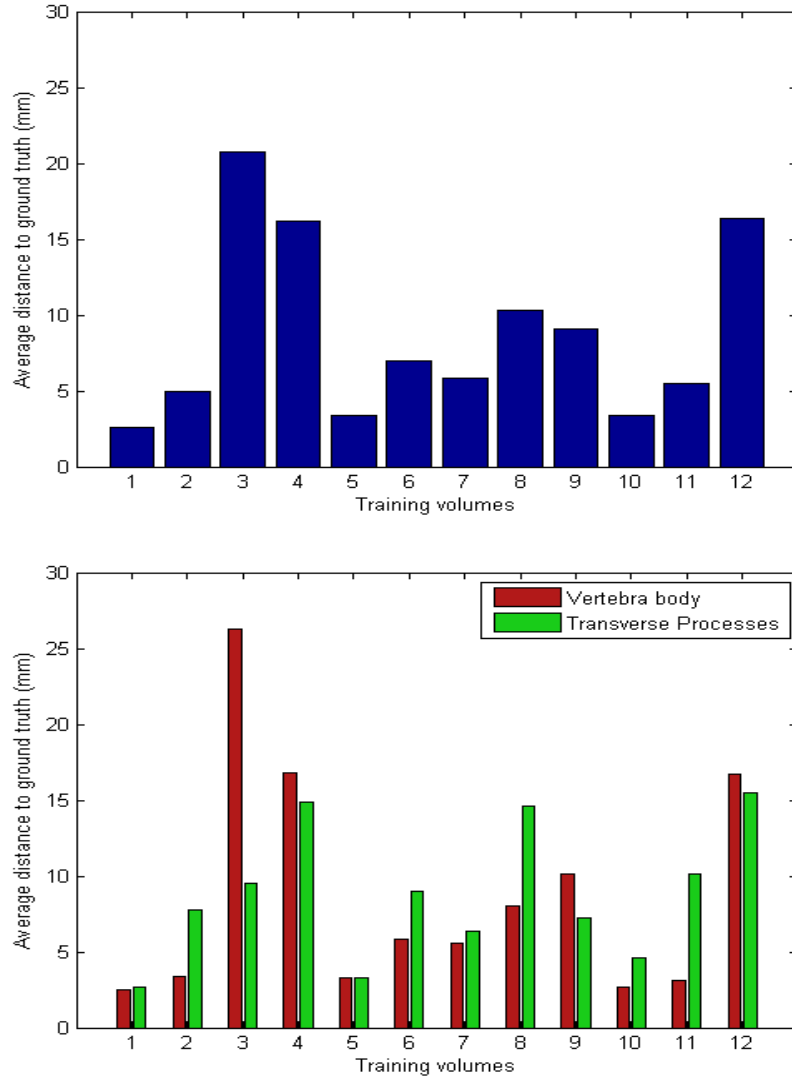


Figure 5.5: Average matching errors based on all 6 anatomical landmarks of the sparse MRF vertebra model (on the top) and average matching errors based on the 4 landmarks of the vertebra body and the 2 landmarks of the transverse processes (on the bottom) by the leave-one-out test with one-dimensional shape index histograms and gradient-based edge descriptors applied on the 12 lumbar vertebra training dataset.

matching, the same type of edge descriptors were extracted along edges between adjacent labels and were compared to the edge descriptors stored in the compatibilites of the model. Although extracting edge descriptors

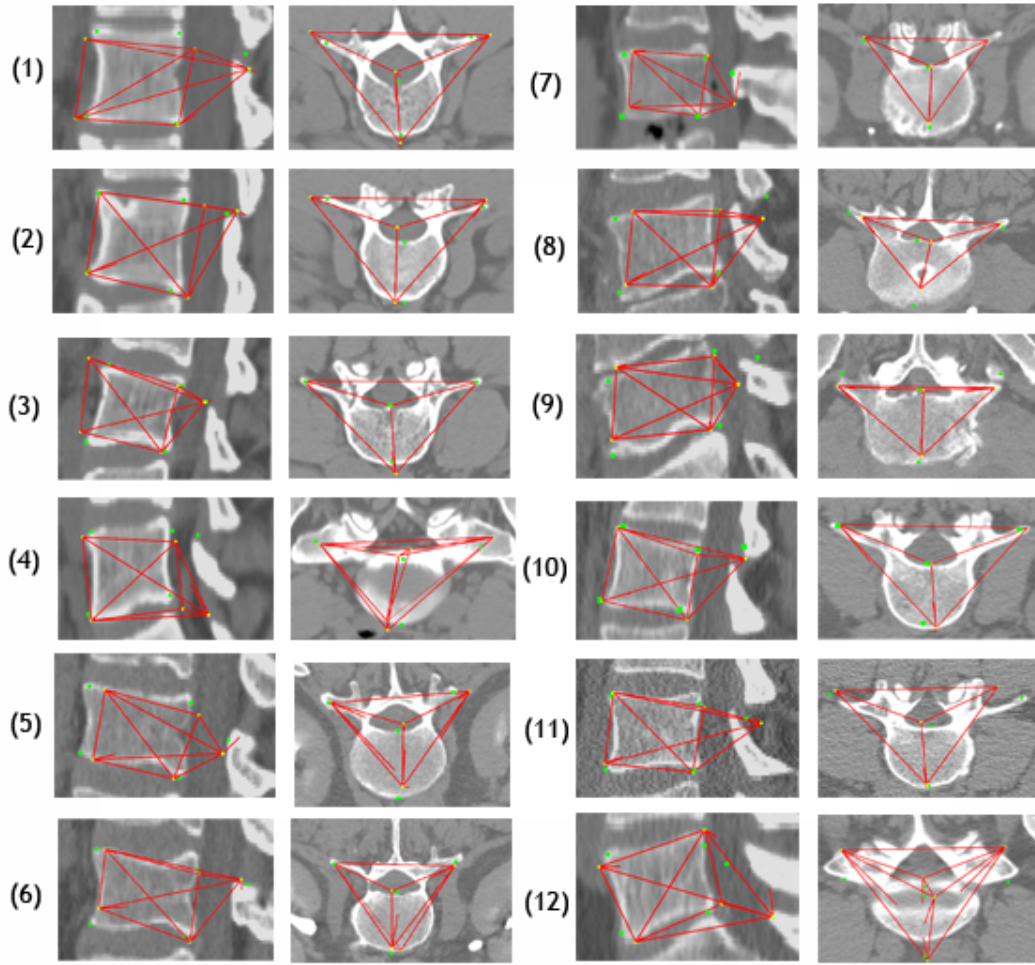


Figure 5.6: MRF matching results by leave-one-out tests using one-dimensional shape index histograms on the 12 lumbar vertebra CT training data.

increased the computational costs for the matching, it delivered additional edge quality information and enhanced the results for the L1-L3 vertebrae. Testing with gradient based edge descriptors and one-dimensional shape index histograms delivered a average matching error of 4.64 mm for the L1-L3 vertebrae and 14.55 mm for the L4-L5 vertebrae where the high error for L4-L5 is due to the high occurrence of rotated vertebra matchings.

5.4 Hardware and Performance

The tests were conducted on a laptop with an Intel Core 2 Duo Processor with 2.20 GHz and 2 GB RAM. A MRF based sparse vertebra localisation run took approximately 1 minute with the one-dimensional shape index histogram descriptor and approximately 4 minutes with the vector spin-image descriptors assuming a maximum of 600 feature point candidates at each landmark modelled by the objects of the sparse MRF vertebra model. Applying gradient-based edge descriptors increased the matching time to approximately 10 minutes.

5.5 Summary

In this thesis a MRF based framework was used for model initialisation of sophisticated vertebra-segmentation algorithms such as AAMs. The MRF framework has been tested with two different local descriptors, with one-dimensional shape index histograms and with vector spin-images. Shape index histogram based matching delivered average matching errors of 5.43 mm and vector spin-image matchings had average matching errors of 5.52 mm for the L1-L3 vertebrae. Testing with one-dimensional shape index histograms and gradient-based edge descriptors the average matching errors could be decreased to 4.64 mm for the L1-L3 vertebrae. The matching performance with maximal 600 feature points per landmark was 1 minute for one-dimensional shape index histogram based matching, 4 minutes for vector spin-images and approximately 10 minutes with additional gradient-based edge descriptors. The tests with gradient-based edge descriptors had poor time performance and its matching accuracy did not differ so much from tests without these descriptors. Thus, considering local descriptors only around landmarks but not along edges were enough for a model initialisation (see figure 5.6).

In summary it can be said that the MRF based sparse structure localisation with shape index histograms and vector spin-images (but without additional edge descriptors) of the L1-L3 vertebrae with the aforementioned MRF based matching accuracy and performance is able to outperform a manual

initialisation of the model of sophisticated full segmentation approaches. The MRF based vertebra model initialisation performed not well (average matching errors approximately 10 mm) for the L4-L5 vertebrae because of rotated vertebra matchings due to the presence of highly similar structures in the vicinity of ground truth landmarks.

Chapter 6

Conclusion and Future Work

6.1 Conclusion

Within this thesis a framework is presented for sparse structure localisation of vertebrae in CT volumetric data. A three-dimensional sparse model of vertebrae shape and appearance is built and matched with target vertebra CT scans where the model-to-target fit is modelled by MRFs. An optimal model-to-target fit is found by solving the MAP-MRF labelling problem.

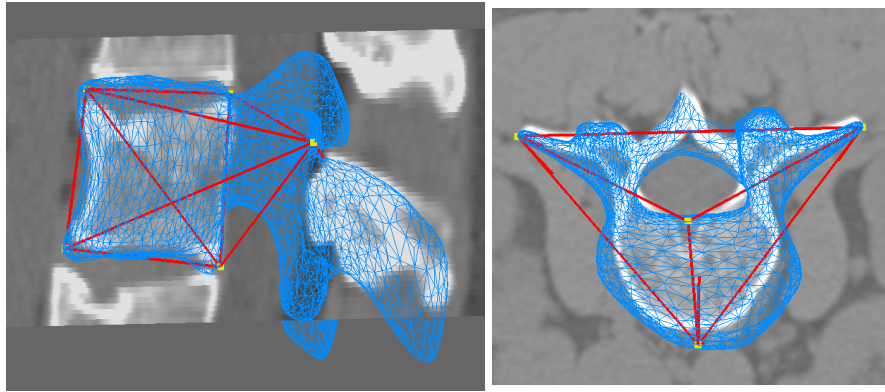


Figure 6.1: The matching result of the MRF framework can be used as an initialisation of more sophisticated segmentation steps such as AAMs.

The sparse model construction involves the selection of local descriptors for anatomically significant landmarks of vertebrae. Furthermore, an adjacency system corresponding to the Markov-property of the MRF graph is set up by Delaunay tetrahedralisation. Euclidean distances and gradient-related

angles along adjacencies set up a geometrical configuration of the model. The built sparse MRF vertebra model is finally matched with candidate locations, called feature points of the target volume. The matching step assigns each object of the sparse MRF vertebra model a feature point with the best fit, taking adjacency relations between them into consideration. Feature points are delivered by a supervised learning algorithm called probabilistic boosting tree (PBT) in the target data. The MAP-MRF labelling problem is solved by the Max-sum approach where a maximum posterior configuration of feature points according to the built sparse MRF vertebra model is delivered.

The sparse MRF vertebra model is built on 12 training volumes of lumbar vertebrae (see section 4.2.1), and the evaluation is established on the same dataset by the leave-one-out strategy. The matching error was measured in millimetres by Euclidean distances between the annotated landmarks at anatomical positions and the matched labels. The MRF based sparse vertebra structure localisation approach resulted for the L1-L3 vertebrae in average matching errors which could be kept in ranges where it could act as an initialisation for subsequent segmentation methods such as AAMs (see figure 6.1). For the L4-L5 vertebrae the MRF approach resulted in high average matching errors (approximately 10 mm) because of rotated vertebra matchings due to the presence of similar anatomical structures (such as the hip) in the vicinity of the modelled landmarks of L4 and L5 vertebrae.

6.2 Future Work

An important future work is to tackle the problem of ambiguous and rotated vertebra matchings, such as in the lower lumbar region (L4 - L5), due to the vicinity of similar bony structures (such as the hips) to the transverse processes. One possibility has been already mentioned in section 5.3 (adding extra landmarks such as the spinous process) which has to be tested and needs further investigation; alternatives should be considered.

Furthermore, the high-level feature extraction technique is planned to be enhanced in order to reduce the number of the resulting feature points per modelled landmark (at the moment it is up to 600). Reducing the number of feature point candidates can accelerate the MRF based vertebra matching.

Enhancing the geometrical configuration is a further point contributing to the future work which could support the stability of the sparse MRF vertebra model. Thus, finding the axis of orientation of each vertebra could provide additional information for the sparse MRF vertebra model and could yield more stability.

More than that, as the MRF approach consists of a combination of different methods according to local descriptors, feature extraction, geometrical setup and a solver, there is a lot of room for exploring new techniques at each of the aforementioned fields. First of all, alternative approaches for the applied MRF solver should be considered to overcome the limitation of the DAG approach of the Max-sum solver, that it can not always be solved to optimality (see section 4.3.4).

Bibliography

- [AKB⁺08] S. Allaire, J.J. Kim, S.L. Breen, D.A. Jaffray, and V. Pekar. Full orientation invariance and improved feature selectivity of 3D SIFT with application to medical image analysis. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 2008. CVPR Workshops 2008*, pages 1–8, 2008.
- [Aus07] Statistik Austria. Ergebnisse der Gesundheitsbefragung 2006/2007. http://www.statistik.at/web_de/dynamic/statistiken/gesundheit/gesundheitsdeterminanten/025390, August 21, 2007. [Online; accessed May 6, 2010].
- [BMP02] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 509–522, 2002.
- [CCS04] D. Cristinacce, T.F. Cootes, and I. Scott. A multi-stage approach to facial feature detection. In *15th British Machine Vision Conference, London, England*, pages 277–286, 2004.
- [CGA10] CGAL. Computational Geometry Algorithms Library. <http://www.cgal.org>, March 22, 2010. [Online; accessed May 6, 2010].
- [Die00] Thomas G. Dietterich. Ensemble Methods in Machine Learning. In *MCS '00: Proceedings of the First International Workshop on Multiple Classifier Systems*, pages 1–15, London, UK, 2000. Springer.

- [DJ95] C. Dorai and A. Jain. Cosmos-a representation scheme for free form surfaces. In *Proc. 5th Intl. Conf. on Computer Vision*, pages 1024–1029, 1995.
- [DKVC98] X. Descombes, F. Kruggel, and D.Y. Von Cramon. Spatio-temporal fMRI analysis using Markov random fields. *IEEE Transactions on Medical Imaging*, 17(6):1028–1039, 1998.
- [DML⁺07] R. Donner, B. Micusik, G. Langs, L. Szumilas, P. Peloschek, K. Friedrich, and H. Bischof. Object Localization Based on Markov Random Fields and Symmetry Interest Points. *Lecture Notes in Computer Science*, 4792:460–468, 2007.
- [DMLB07] R. Donner, B. Micušik, G. Langs, and H. Bischof. Sparse MRF appearance models for fast anatomical structure localisation. In *British Machine Vision Conference*, volume 1, 2007.
- [DWBL09] R. Donner, H. Wildenauer, H. Bischof, and G. Langs. Weakly supervised group-wise model learning based on discrete optimization. In *MICCAI '09: Proceedings of the 12th International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 860–868, Berlin, Heidelberg, 2009. Springer-Verlag.
- [EPS⁺08] O. Ecabert, J. Peters, H. Schramm, C. Lorenz, J. Von Berg, MJ Walker, M. Vembar, ME Olszewski, K. Subramanyan, G. Lavi, et al. Automatic model-based segmentation of the heart in CT images. *IEEE Transactions on Medical Imaging*, 27(9):1189–1201, 2008.
- [GG84] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Recognition and Machine Intelligence*, 6:721–741, 1984.
- [GO04] J.E. Goodman and J. O'Rourke. *Handbook of discrete and computational geometry*. Chapman & Hall/CRC, 2004.

- [HB04] M. Hahn and T. Beth. Balloon based vertebra separation in CT images. In *Proceedings of the 17th IEEE Symposium on Computer-Based Medical Systems (CBMS 2004)*, volume 1063, pages 310–315, 2004.
- [HBG⁺84] IL Herlin, D. Béréziat, G. Giraudon, C. Nguyen, and C. Graffigne. Segmentation of echocardiographic images with Markov random fields. Technical report, Institut national de recherche en informatique et en automatique, 1984.
- [HKK⁺97] K. Held, E.R. Kops, B.J. Krause, WM Wells, R. Kikinis, and H.W. Muller-Gartner. Markov random field segmentation of brain MR images. *IEEE Transactions on Medical Imaging*, 16(6):878–886, 1997.
- [HL07] N. Hosten and T. Liebig. *Computertomographie von Kopf und Wirbelsäule*. Thieme, 2 edition, 2007.
- [HLLS01] G. Hetzel, B. Leibe, P. Levi, and B. Schiele. 3D object recognition from range images using local feature histograms. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2. IEEE Computer Society; 1999, 2001.
- [JH99] A.E. Johnson and M. Hebert. Using spin images for efficient object recognition in cluttered 3D scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(5):433–449, 1999.
- [KHB06] S. Koompaiojn, K.A. Hua, and C. Bhadrakom. Automatic classification system for lumbar spine x-ray images. In *19th IEEE International Symposium on Computer-Based Medical Systems, 2006. CBMS 2006*, pages 213–218, 2006.
- [KS80] R. Kindermann and J.L. Snell. Markov Random Fields and their Applications. First book of the AMS soft-cover series in Contemporary Mathematics. *American Mathematical Society*, 12, 1980.
- [KvD92] J.J. Koenderink and A.J. van Doorn. Surface shape and curvature scales. *Image and vision computing*, 10(8):557–564, 1992.

- [KWL⁺08] T. Klinder, R. Wolz, C. Lorenz, A. Franz, and J. Ostermann. Spine segmentation using articulated shape models. In *Proceedings of the 11th international conference on Medical Image Computing and Computer-Assisted Intervention-Part I*, pages 227–234. Springer, 2008.
- [KWTM03] G. Kindlmann, R. Whitaker, T. Tasdizen, and T. Moller. Curvature-based transfer functions for direct volume rendering: Methods and applications. In *Proceedings of the 14th IEEE Visualization 2003 (VIS'03)*, page 67. IEEE Computer Society, 2003.
- [Lip06] Lippert. *Lehrbuch Anatomie*. Urban & Fischer, 7th edition, 2006.
- [LO06] H. Ling and K. Okada. Diffusion Distance for Histogram Comparison. In *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Volume 1*, pages 246–253. IEEE Computer Society, 2006.
- [Low99] D.G. Lowe. Object recognition from local scale-invariant features. In *International Conference on Computer Vision*, volume 2, pages 1150–1157, 1999.
- [Min10] University of Minnesota. Anatomical Variations of Vertebrae. <http://msjensen.cehd.umn.edu/imagebank/Bones/vertebra2.gif>, April 2, 2010. [Online; accessed May 6, 2010].
- [MNP06] P. Mohapatra, P. Nanda, and S. Panda. Color image segmentation using MRF model and simulated annealing. *Proceedings of Soft Computing Technique for Engineering Applications (SCT 2006)*, 2006.
- [MS05] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1615–1630, 2005.
- [NA08] M.S. Nixon and A.S. Aguado. *Feature extraction and image processing*. Academic Press, second edition, 2008.

- [Nae07] B. Naegel. Using mathematical morphology for the anatomical labeling of vertebrae from 3D CT-scan images. *Computerized Medical Imaging and Graphics*, 31(3):141–156, 2007.
- [Nas97] C. Nastar. The Image Shape Spectrum for Image Retrieval. Technical report, Institut National De Recherche En Informatique Et En Automatique, INRIA Paris - Rocquencourt, 1997.
- [Per98] P. Perez. Markov random fields and images. *CWI Quarterly*, 11(4):413–437, 1998.
- [RCA06] M.G. Roberts, T.F. Cootes, and J.E. Adams. Automatic segmentation of lumbar vertebrae on digitised radiographs using linked active appearance models. In *MIUA*, volume 6, pages 120–124, 2006.
- [RKD06] M. Reiser, F.P. Kuhn, and J. Debus. *Duale Reihe Radiologie*. Thieme, 2 edition, 2006.
- [Rob01] A. Roberts. Curvature attributes and their application to 3D interpreted horizons. *First Break*, 19(2):85–100, 2001.
- [RTG00] Y. Rubner, C. Tomasi, and L.J. Guibas. The earth mover’s distance as a metric for image retrieval. *International Journal of Computer Vision*, 40(2):99–121, 2000.
- [Rub99] Y. Rubner. Code for the Earth Movers Distance (EMD). <http://ai.stanford.edu/~rubner/emd/default.htm>, January 29, 1999. [Online; accessed May 6, 2010].
- [RW08] B. Roeschies and S. Winter. Detection of vertebrae in CT slices by bunch graph matching. In *Processing European Congress for Medical and Biomedical Engineering*, volume 22, pages 2575–2578, 2008.
- [Sch76] M. Schlesinger. Sintaksicheskiy analiz dvumernykh zritel’nykh signalov v usloviyakh pomekh (syntactic analysis of two-dimensional visual signals in noisy conditions). *Kibernetika*, 4:113–130, 1976.

- [Sta95] Z. Li Stan. *Markov Random Field Modeling in the Computer Vision*. Springer, 1995.
- [TBA⁺09] P.A. Tresadern, H. Bhaskar, S.A. Adeshina, C.J. Taylor, and T.F. Cootes. Combining Local and Global Shape Models for Deformable Object Matching. In *British Machine Vision Conference 2009*, 2009.
- [Tu05] Z. Tu. Probabilistic boosting-tree: Learning discriminative models for classification, recognition, and clustering. In *Proceeding of ICCV*, volume 3, 2005.
- [Uni10] Spine Universe. Vertebral Column. <http://www.spineuniverse.com/anatomy/vertebral-column>, March 29, 2010. [Online; accessed May 6, 2010].
- [VJ02] P. Viola and M. Jones. Robust real-time object detection. *International Journal of Computer Vision*, 57(2):137–154, 2002.
- [VRV10] VRVis. Zentrum für Virtual Reality und Visualisierung Forschungs-GmbH. <http://www.vrvis.at>, April 24, 2010. [Online; accessed May 6, 2010].
- [WC07] J. Wu and A.C.S. Chung. Markov Random Field Energy Minimization via Iterated Cross Entropy with Partition Strategy. In *IEEE International Conference on Acoustics, Speech and Signal Processing, 2007. ICASSP 2007*, volume 1, 2007.
- [Wer05] T. Werner. A Linear Programming Approach to Max-sum Problem: A Review. Technical report, Center for Machine Perception, Department of Cybernetics Faculty of Electrical Engineering, Czech Technical University, 2005.
- [Wer07] T. Werner. A Linear Programming Approach to Max-sum Problem: A Review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(7):1165–1179, 2007.

- [Wer09] T. Werner. Max-Sum (Labeling) Problem Page. <http://cmp.felk.cvut.cz/~werner/software/maxsum/>, January 15, 2009. [Online; accessed May 6, 2010].
- [XD08] L. Xu and H.Q. Dinh. A local descriptor for finding corresponding points in vector fields. In *IAPR International Conference on Pattern Recognition (ICPR)*, pages 1–4, 2008.
- [ZBG⁺08] Y. Zheng, A. Barbu, B. Georgescu, M. Scheuering, and D. Comaniciu. Four-chamber heart modeling and automatic segmentation for 3D cardiac CT volumes using marginal space learning and steerable features. *IEEE Transactions on Medical Imaging*, 27(11):1668–1681, 2008.