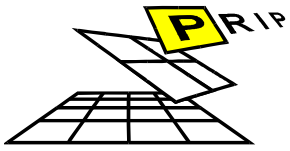
PRIP-TR-120

# Automatic Model Generation for Sparse MRF Appearance Models using Minimum Description Length

*Eva Dittrich*

**P**attern
**R**ecognition &
**I**mage
**P**rocessing
Group




Institute of
Computer Graphics and
Algorithms

Technical Report

Pattern Recognition and Image Processing Group
Institute of Computer Aided Automation
Vienna University of Technology
Favoritenstr. 9/1832
A-1040 Vienna AUSTRIA
Phone: +43 (1) 58801-18351
Fax: +43 (1) 58801-18392
E-mail: eva.dittrich@student.tuwien.ac.at
URL: http://www.prip.tuwien.ac.at/

PRIP-TR-120 January 6, 2009

# Automatic Model Generation for Sparse MRF Appearance Models using Minimum Description Length

*Eva Dittrich*

**Abstract**

This report explores the combination of the Minimum Description Length (MDL) approach and Sparse Markov Random Fields Appearance Models (SAMs). SAMs are a method to locate a structure that is learnt from annotated training data in a new and unseen image. However, to achieve this result it is necessary to provide the SAMs with manual annotations of the images (landmarks) for a large set of training examples, which is a time consuming and error-prone requirement. The goal of this work is to become independent from manual annotations and to obtain the annotations automatically by using an MDL based approach. We report experimental results for different types of data (synthetic data, 2D X-rays and 3D CTs) and the method was modified to reach the best possible results for each of them. The resulting approach allows to construct SAMs in a fully automated fashion.

# Contents

# Chapter 1

# Introduction

This report documents the effort to combine two powerful principles: *Minimum Description Length (MDL)* based shape model building and *Sparse MRF Appearance Models (SAMs)* [1]. The former approach was modified and tested extensively to automatically build sparse shape models for medical X-ray images. Based on these models, SAMs were used to locate the structures in new (unknown) pictures.

## 1.1 Motivation

Automatically locating geometric structures in images is a challenge in computer vision and at the same time important in medical science [2] Achieving an automatic localization of anatomical structures in medical images provides a crucial initialization step to segmentation methods like ASMs [3], AAMs [4] or graph-cuts [5] [6].

Sparse MRF Appearance Models (SAMs; see Ch. 2.2) enable an automatic mapping of a model learnt from a set of training images to a new, unseen image. However, this process is based on a preassigned annotation of training images from which a statistical model of shape and appearance is built. The annotation consists of a selection of interest points - or landmarks - (see Ch. 2.1) that are detected in each image. For a set of these interest points correspondences have to be established across all training images, or volumes. These points are called *landmarks* and allow for the consistent identification of positions in all images (e.g. landmark number one always marks the left index finger tip in each picture).

At this point a problem arises: Currently the annotation is done by manually clicking on landmarks (which is problematic and extensive in 3D; see Ch. 1.2), hence there is a need for an automatic generation of these annotations.

## 1.2 Fundamental Concepts and Terms

Before describing the method in detail, it is necessary to explain a few fundamental concepts and terms.

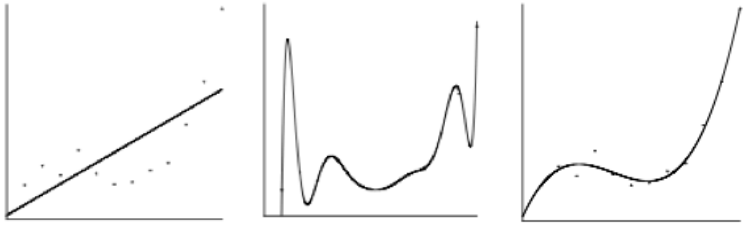
**Shapes and Landmarks.** A shape is defined as a collection of corresponding border points or the characteristic surface configuration of an object: an outline or a contour [7]. It can further be seen as sets of landmarks. Landmarks (LMs) are salient points of an object or structure. They can be used as points of correspondence on each image (e.g. landmark number one always marks the left index finger tip in each picture). According to [7] there are three kinds of LMs:

- Anatomical LMs - points that have some biological meaning in organisms and are marked by an expert (e.g. a doctor).
- Mathematical LMs - points that arise out of some mathematical or geometrical calculation (e.g. extremum points).
- Pseudo-LMs - artificially created points along a border or between other LMs.

Until recently, the state of the art for acquiring LMs for the training of SAMs has been to manually place them on the pictures (i.e., anatomical LMs)[1]. This procedure costs time and effort in 2D and the results are partly subjective and therefore sub-optimal [8]. When expanding the dimension to 3D or even 4D (including time as fourth dimension) it becomes unfeasible because of the circumstances to navigate within a 3D space in order to annotate volumes. It should also be kept in mind that annotations made by human experts include the possibility of errors which make it even harder to achieve a proper standard of reference that is used to evaluate the results of medical image processing, and can introduce a significant amount of noise in the resulting models.

**Hypotheses and Models.** A (point) hypothesis represents a single probability distribution or function [9], e.g. the polynomial $4x^3 - 9x^2 + x - 3$ to sample a set of coordinate points. A model refers to a set of hypotheses, e.g. the set of all polynomials of third-degree. Speaking of sampling points, there are three ways of retrieving a model for a given data set. Consider the three parts of Fig. 1.1: on the left side (Fig. 1.1(a)) the function does not reveal any information about the points so it does not fit the data points good enough. In this case, the points are obviously sampled in a too simple way,

which is described as **underfitting**. In contrast to this, Fig. 1.1(b) shows an exact interpolation of the data points. Although in this specific example the points are sampled perfectly, it will immediately become insufficient, as soon as another data point is added lying clearly outside of the function graph. This phenomenon is called **overfitting**. Finding a **trade-off** between a too simple and too complex models leads to a much better description and prediction of the data (see Fig. 1.1(c)).



(a) (b) (c)

Figure 1.1: Examples for (a) underfitting, (b) overfitting and (c) trade-off polynomials to sample a dataset. The data points are not only taken from a polynomial but also perturbed by noise. [9]

By combining the information from the annotations of the training images a statistical model of shape and appearance is built. An ideal model has several *ideal* properties: [10]

- Generalization: the model represents any instance of the object (images), not only those from the training process.
- Specificity: the model describes only valid instances of the object.
- Compactness: the variation in the model is described with as few parameters as possible.

The MDL principle (see Ch. 2.3) is an approach to implement these demands. It finds regularities in data sets and thus compresses the data. See [11] for an extensive explanation of this method. Having created a representative model with help of the MDL principle, SAMs can be applied to automatically locate anatomic structures in test images. This means that no tedious, time-consuming and manual selection of interest points is necessary, allowing to apply SAMs more efficiently to new modalities.

The remaining report is structured as follows: Ch. 2 presents the methods that are used in this work. It describes different kinds of interest points (Ch. 2.1), Sparse MRF Appearance Models (Ch. 2.2), the MDL principle

(Ch. 2.3) and finally Principal Component Analysis (Ch. 2.4). In Ch. 3 we explain the implementation of these methods. The experiments including their setup as well as results are given in Ch. 4 and the final conclusion can be found in Ch. 5.

# Chapter 2

# Methods

## 2.1 Interest Points

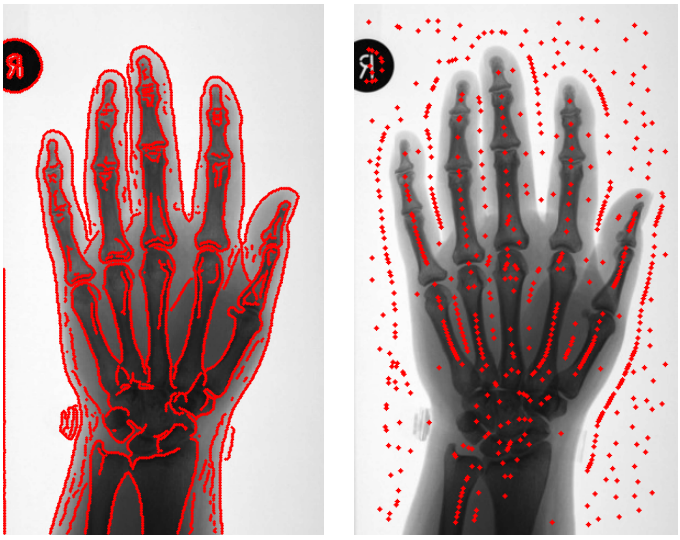
An interest point is a point in an image that can be described by a clear mathematically definition, a particular position in the image or a surrounding image structure. Only such interest points can be computed with a high degree of reproducibility and are therefore locally and globally stable.

### 2.1.1 Symmetry Interest Points

In medical images there are various anatomical structures (like bones or veins) that are of interest to medical experts and reveal a shape with a certain degree of symmetry regarding at least one axis. This (local) symmetry cannot only be found in normal X-ray images but also in 2D slices of a 3D data set (like an MRI) [1].

Other interest points which are detected by popular approaches like the Canny edge detector (which retrieves edges in images that the points are taken from; introduced in [12]) and the Difference of Gaussians (DoG) are used for edges or corner-like structures, but they do not deal with local symmetry (a comparison between the Canny edge detector and symmetry interest points can be seen in Fig. 2.1(a) and Fig. 2.1(b)).

To retrieve interest points of local symmetry, the Gradient Vector Flow (GVF) field is used (for more details see [13]). It increases the capture range of active contours and is even able to detect weak structures while preserving a robustness to a high amount of noise in the image. The GVF of an image $I$ can either be computed from a binary edge map or directly from the gray level image as $GVF(I) = G = u + i * v$, resulting in the complex matrix and field $G$. The field magnitude $|G|$ is largest in areas of high image gradient. The field will point towards or away from the local symmetry center of the



(a) (b)

Figure 2.1: Interest points retrieved by (a) Canny edge detector and (b) symmetry interest points as local minima from GVF.

structure. Therefore the symmetry interest points are defined as the local minima of $|G|$ [1].

## 2.2 Sparse Markov Random Field Appearance Models

Sparse Markov Random Field Appearance Models (SAMs) are an approach for fast and reliable structure detection as well as segmentation in medical images.

Segmentation approaches like Active Shape Models (ASMs), Active Appearance Models (AAMs), Active Features Models (AFMs), graph-cuts and snakes require a reasonable initialization of the model search: For ASMs and AAMs the images (or objects of interest) need to be overlapped, graph-cuts require manually placed points inside and outside of the object of interest, whereas snakes rely upon spatial constraints for a precise description of the object of interest [1].

**Markov Random Fields** MRFs incorporate graphs with $M$ nodes and $N$ fields or labels for each node. The nodes represent objects and the labels provide associated qualities. The labels of two adjacent nodes are fully con-

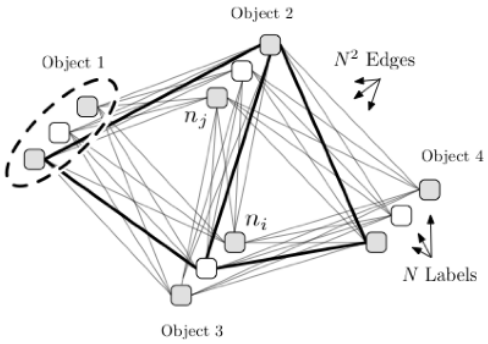
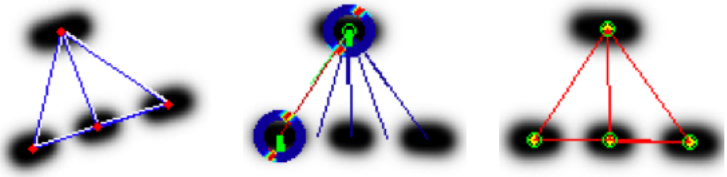



Figure 2.2: The MRF graph consisting of $M = 4$ objects with $N = 3$ labels each. The sub-graph symbolized by thick lines is the goal which is equal to the maximization of the sum of label and edge qualities.

nected by $N^2$ edges, each of them being assigned a weight to encode quality. To identify the adjacency of two objects, there is an additional graph $\mathcal{A}$ with $A$ edges, representing the adjacencies between objects. Using the MAX-SUM problem solver, the goal can be achieved which is to select one label for each object to maximize the sum of label and edge qualities which results in a sub-graph (see Fig. 2.2).

In the MAX-SUM problem of second order a sum of bivariate functions of discrete variables is maximized. The result of the MAX-SUM problem is finding a configuration for a Gibbs distribution with maximal probability which means the same as finding a maximum posterior configuration of an MRF with discrete variables [1]. Additionally it permits to define several labels while still keeping the processing time within reasonable bounds [1]. Apart from the MAX-SUM solver there are other efforts to solve the multi-label problem for MRF, e.g. second order cone programming, sequential tree-reweighted max-product message passing or belief propagation methods [1]. However, none of them (including the MAX-SUM problem) solves the problem exactly because it is NP-hard [1]. In the case of the graph being a tree, the global optimum can be found – otherwise the MAX-SUM problem takes several approximations into account to reach a possibly optimal solution [1].

**Sparse Appearance Models** SAMs obtain information from images using local descriptors around interest points and along edges between these points. There is no PCA model used in order to obviate the need for a large number of training sets and the global character of PCA-based models. With help of a Delaunay triangulation of the model's interest points (symmetry in-



(a) (b) (c)

Figure 2.3: Applying Sparse MRF Appearance Models to synthetic images. (a) The selected model points including the connecting edges between adjacent points. (b) Building the MRF: the visualization during the computation of statistical models (the edges' lengths, relative angles and local descriptors). (c) The resulting image, predicting and mapping the selected points to an unseen image.

terest points, see Ch. 2.1.1), the shape model is constructed. In addition, statistical models of the edges' lengths, relative angles and local descriptors are computed, which altogether result in models that are locally deformable and rotation invariant [1].

In each of the $n$ model images, $M$ interest points are manually selected describing the anatomical structure to be found. One model image is used to compute the Delaunay triangulation of its $M$ interest points, which results in adjacencies of model points that generate or can be described by the graph $\mathcal{A}$. A solution $\mathbf{S}$ corresponds to a mapping of the $M$ selected model points (representing the objects of the MRF graph) to a subset of the $N$ target interest points (representing the labels) [1].

**Constructing Sparse Random Markov Field Appearance Models** Using a given SAM and a target image, the MRF is applied to describe the confidences that a model point or edge should be matched to a certain interest point or edge in the target image [1]. Due to the fact that in this case a maximization problem has to be solved, all confidences or qualities are located within $[-\infty, 0]$. The descriptor distances are normalized to having a maximum of 0 and a median of $-1$, whereas the length and angle confidences are in the interval $[-1, 0]$ [1].

The quality of a (model point, target point)-match is equal to the negative distance between the model point descriptor $\overline{D}_m$ and the local target descriptor. To achieve all two-way distances between model and potential target it is necessary to compute correspondences which result in the $M \times N$ matrix $\mathbf{C}$. The edges' qualities in the model are stored in $\mathbf{E}$, where the quality of an edge $e$ between two labels $n_i, n_j$ in $\mathbf{E}$ is computed by comparing its length

$l_e$ and relative angles $\beta_{e1}, \beta_{e2}$ to the corresponding Gaussian distributions of the model edge (see [1] for more detail).

It is possible that in one location of a medical structure in the target image no interest point can be found where the model would expect one. Therefore it is important to include the chance of leaving out a model point which is enabled by adding one artificial target interest (dummy) point [1].

## 2.3 Minimum Description Length

The MDL principle was first introduced by Jorma Rissanen in 1978 [14]. In contrast to Bayesian inference which is formulated in a probabilistic framework, the MDL principle can be viewed as purely data driven [15]. Another closely related approach called Minimum Message Length (MML) was introduced in 1968 by Wallace and Boulton and is based on making a priori assumptions as well. MML had been misleadingly confused with MDL not only once (for more details see [16]).

The idea of the MDL approach was to formalize *Occam's Razor* which basically states that the simpler the solution to a problem, the better it is. It means that among a set of equally competing hypotheses explaining a problem, the one with the fewest assumptions and requirements shall be chosen. Therefore, interpreting the simplest solution as the one with the shortest description, the best hypothesis is the most compressed one [15]. Setting *compressing data* equal to *learning* it is clear that the more the data can be compressed, the more we *learned* about the data.

Occam's Razor led to the following concept: the more regularities there are in a data set, the more the data can be compressed and therefore described by fewer symbols [9]. The length of the shortest possible description of data is its *Kolmogorov complexity* [15]. In order to find the hypothesis that compresses the data the most, one must choose a computer language to write a program of shortest length which describes the data. The choice of the computer language is not important because the shortest program in either language varies only by a constant which can be disregarded in case of sufficient long data. As it is shown in [9] no computer program can be found that returns the shortest program which represents the data. Furthermore, in practice there are only small data sets and consequently the constant factor carries authority.

This ideal principle of MDL is not applicable in practice [9]. Rather description methods should be used that are less expressive than general-purpose computer languages [9] but limiting enough to describe the length of the shortest representation of the data as well as general enough to compress

as many regular sequences as possible.

An advantage of MDL is that when being used for model selection, it avoids overfitting of the data. It provides a trade-off between too simple and too complex models.

There are two versions of the practical MDL principle: the *two-part code* and the *one-part code*. In the former version, to achieve the best hypothesis $H_i$ for describing the data $D$, the sum in Eq. (2.1) must be minimized, where $L(H)$ is the length of the hypothesis and $L(D|H)$ is the length of the description of the data when coded by the hypothesis.

$$L(H) + L(D|H) \tag{2.1}$$

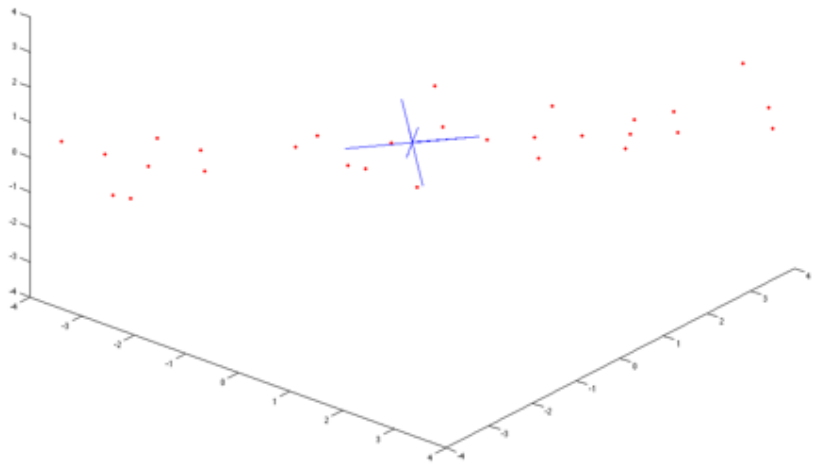
It is not enough to minimize only one summand, because when finding a good fit (small $L(D|H)$) it is usually a complex hypothesis (with a large $L(H)$), whereas a simple hypothesis (small $L(H)$) has a rather bad fit (large $L(D|H)$) [9].

In the *one-part code* version of the MDL principle (also called *refined MDL* in [9]), the aim is not to find a single hypothesis, but rather a full model $\mathcal{H}$ (which means a set of hypotheses) for encoding the data. The condition is not split in two parts, but only a single *one-part code* with lengths $\overline{L}(D|\mathcal{H})$, where this codelength will also be small whenever there is a hypothesis $H \in \mathcal{H}$ that fits the data well (small $L(D|H)$). In refined MDL there is a second concept called *parametric complexity* $\mathbf{COMP}(\mathcal{H})$ which indicates a model's ability to fit random data [9]. Model selection based on refined MDL is still a trade-off between the goodness of fit (described by $L(D|\widehat{H})$, where $\widehat{H}$ is the distribution in $\mathcal{H}$ which minimizes the codelength) and the complexity term $\mathbf{COMP}(\mathcal{H})$.

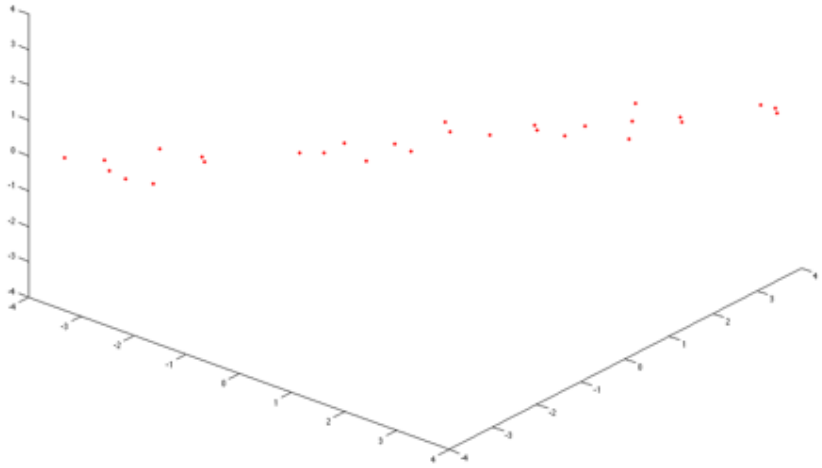
## 2.4 Principal Component Analysis

Principal Component Analysis (PCA) is a statistical technique for dimensionality reduction that can be used in areas like face recognition and image compression.

Considering a 3D scatterplot (as shown in Fig. 2.4(a)) with the data $X = (x\ y\ z)$ as an $n \times 3$-matrix. $n$ is the number of points and $x$, $y$ and $z$ are columns representing the coordinates $x = (x_1, ..., x_n)^T$, $y = (y_1, ..., y_n)^T$ and $z = (z_1, ..., z_n)^T$. The first step is to normalize the data points $x = x_i - \overline{x}$ with $i \in \{1, ..., n\}$ (similarly for $y$ and $z$) where $\overline{x}$ is the mean of the $x$-coordinates ($\overline{y}$ and $\overline{z}$ likewise). This guarantees that the scatterplot is centered at the



(a)



(b)

Figure 2.4: 2.4(a) The normalized 3D scatterplot with the eigenvectors (blue lines). 2.4(b) The resulting subspace when projecting the data onto the two eigenvectors with the largest variance.

point of origin. The next step is to calculate the covariance matrix c:

$$c = \begin{pmatrix} cov(x,x) & cov(x,y) & cov(x,z) \\ cov(y,x) & cov(y,y) & cov(y,z) \\ cov(z,x) & cov(z,y) & cov(z,z) \end{pmatrix}$$

Having calculated the eigenvectors and eigenvalues of $c$, the rotated basis $F = (e_1, ..., e_m)$ is constructed where $e_i \quad (i \in \{1, ..., m\})$ are the eigenvectors sorted by the corresponding eigenvalue in descending order (in this case $m = 3$ because there are 3 dimensions and therefore 3 eigenvectors; the eigenvectors can also be seen in Fig. 2.4(a)). In case only the first $p$ eigenvectors out of $e_1, ..., e_p, ..., e_m$ (leaving out the last $m - p$ ones with smallest eigenvalues) are selected, the other dimensions are disregarded and the dimensionality of the data is therefore reduced. See Fig. 2.4(b) for an example where the data points are projected onto the first two eigenvectors.

The MDL principle uses PCA to build a statistical model of shape variation. In order to achieve this model, all sets of landmarks (one set representing the salient landmarks of a picture) are aligned to exclude rotation, translation and scaling variation. Then shape vectors are generated where each vector includes the coordinates of the landmarks of a picture. After this step PCA is performed on the shape vectors, resulting in a statistical shape model. This shape model contains the mean shape as well as a set of modes (the eigenvectors of the covariance matrix which are plausible deformations of the shape).

PCA is also used in several other parts of medical image processing: ASMs, AAMs, PCA-SIFT, structural tensor, etc. However, these are not part of this project and therefore not explained any further.

# Chapter 3

# Employing MDL-based annotations for SAMs

This section deals with the concrete implementation of the MDL principle as well as the SAMs approach. Sec.3.1 describes the implementation of the interest points that are used in this work. The functionality of the MDL principle is described in Sec.3.2, the application of models in SAMs is explained in Sec.3.3.

## 3.1 Interest Points

During the progress of this project several different kinds of interest points were needed to meet different demands. The main types were symmetry interest points (Ch. 2.1.1) and points retrieved from the Canny edge detector (Ch. 2.1.1).

Depending on the type of image, different types of interest points turned out to be the best solution (see Tab. 3.1).

| image type | interest points |
|---|---|
| 2D synthetic images | Canny edge detector |
| 2D hand X-ray images | symmetry interest points |
| 3D hand CT images | symmetry interest points |

Table 3.1: The different kinds of images and the best selection of landmarks for them.

## 3.2 MDL based Correspondences of Landmarks

The implementation of the MDL principle by [8] was used as starting point in this project. It starts after the preparation of the images that delivers the interest points for the images (for details see Ch. 3.3. In addition, a local descriptor is computed for each landmark (e.g. SIFT, steerable filters). Among all retrieved interest points (which will be treated as landmark candidates) $m$ are selected out of one example image $I_i$ ($i \in 1, ..., n$) and their indices (referring to the actual coordinates) are stored in an $m \times n$ matrix *genome* which includes $m$ interest points' indices for each of the $n$ images. A subsequent initial match finds the corresponding (or in this first step: the *closest*) landmarks in all other images $I_1, ..., I_n$.

After this initial step the optimization starts. $n - 1$ images are used to build a model for shape as well as the local textures around each of the landmarks. The shape model is constructed out of the current landmarks in each image. This model is then fitted to the remaining image $I_k$ and the resulting deviation from the shape model is calculated. Each landmark is replaced by the candidate point with the smallest cost, which emerges as a combination of texture and shape costs. The shape costs are however not only the distances to the shape model but also the distances to what the model is able to reconstruct (e.g. considering a strong variation in one direction of the shape model that is built out of $n-1$ images: In this case it is "cheaper" to choose candidate points out of $I_k$ that differ within the same orientation). According to [8] it is shown that although this is no brute force minimization of the MDL criterion, the model is minimized anyway.

The optimization procedure runs until convergence or until a maximum number of iterations is reached, yielding the best *genome* (the $m \times n$-matrix, including all $m$ corresponding landmarks for each image $I_i, i \in \{1, ..., n\}$).

## 3.3 Using SAMs with the automatically obtained models

In this project the Sparse MRF Appearance Model implementation by [1] was used as a basis. First of all the parameters for the SAMs as well as for the MDL approach are loaded from an external settings file. Then the process continues by preparing all $n$ images and computing interest points, local descriptors as well as several more attributes for each image $I_i$. All data is stored in a structure including the image data for all $n$ images, the calculated interest points, features for each point and other attributes for each image $I_i$. Afterwards the results from the MDL step are loaded which

are stored in an $m \times n$ matrix containing $m$ corresponding interest points for all $n$ images. Having created a representative model with help of the MDL principle, Sparse MRF Appearance models can be applied to automatically locate anatomic structures in test images which means that no tedious, time-consuming and manual selection of interest points is necessary.

In the original version each image was loaded and the user had to manually select several landmarks on each one, paying attention to click on the landmarks in the same order in each picture. This step was eliminated because the MDL principle already provides the corresponding landmarks for all pictures. Then the graph representing the model points' neighbourhoods is computed as well as the adjacencies between the model points (nodes) and angles between model edges (concatenating two model points) which results in a Delaunay triangulation of the model points.

During the final step of the process a leave-one-out test with all training images is run. During this evaluation $n - 1$ images are used to compute the SAMs and another one that will be treated as unseen and therefore *target* image $I_t$. The trained knowledge is now mapped to the unknown image (see. Fig 2.3). The result of the SAMs is a set of corresponding interest points for the target image $I_t$.

# Chapter 4

# Experiments

For evaluation three different kinds of data were used: synthetic 2D images, X-rays of human hands and CTs of human hands. The MDL principle was evaluated for both 2D datasets using a set of images and interest points which represented the ground truth (these images were annotated manually). The median and mean error of the distance of resulting landmarks to ground truth landmarks were used to measure the performance of the approach for 2D objects.

In addition to the error rates, the results were also visualized. At this point it is important to compare the initialization with the optimization result. In the images after the initizalization step, the landmarks result from the computation of the shortest distances, i.e. the correspondence to a point in the first image is initialized by choosing the closest point in the second image. Descriptors or other features are not regarded in this case, therefore the landmarks seem to be arranged similarly in all images (regarding the distances between the landmarks), but do not cover the same areas or objects. The visualization of the MDL result (after the optimization) reveals corresponding landmarks, i. e. landmarks that represent the same area in all pictures.

When assessing the initialization and optimization result for the 3D dataset - where currently no expert ground truth is available - it is necessary to decide whether or not the images seem plausible. Holes, implausible topology changes, and distorted structures are an indication of erroneous correspondences.

## 4.1 Setup

### 4.1.1 Synthetic 2D Images

20 synthetic images were created showing a rectangle which grows in width from image 1 to image 20 (see Fig. 6.2). These simple images were used to get acquainted with the MDL procedure and the parameters governing its behavior. It was tested whether MDL is able to find reasonable landmarks, i. e. whether the landmarks representing corners are consistently selected in all images.

The Canny edge detector was applied to the images to retrieve interest points along edges and corners and MDL was used to find correspondences between these points. The results of the MDL approach for this dataset was not used for the SAMs because it was rather meant to test the MDL behaviour.

### 4.1.2 Hand X-Rays

20 X-ray images of human hands were used to evaluate the MDL results and SAM behavior. The images had a size of about 300 $\times$ 500 pixels each. To extend the data set, not only 10 X-rays of human left hands were used but also 10 X-rays of human right hands were mirrored and treated as left hands. The images captured only healthy hands. In each image around 300 symmetry interest points were found.

Initial tests showed that the MDL principle works best with about 20 to 30 selected points equally distributed over each image. The hand images revealed much better results when having selected symmetry interest points instead of points retrieved by the Canny edge detector because symmetry interest points Therefore the former are used for tests and experiments with these kind of images.

The goal of the MDL step was to find 30 corresponding points in each of the 20 images. To evaluate the results, the mean and median deviation of the model points to ground truth data was calculated. The MDL result was used for the SAMs to replace the initial step of manually selecting landmarks out of a set of interest points. SAMs were then applied to use the model points retrieved by the MDL procedure and find them in a new and unseen target image. Finally the mean and median error between the ground truth and by the SAMs retrieved landmarks was calculated.

### 4.1.3 Hand CT images

The third and most complex dataset consisted of 6 CT volumes of human hands. Symmetry interest points were used as model points in this case. For each volume between 1650 and 1850 symmetry interest points were found. Several experiments showed that good MDL results could be achieved by using only 25 landmarks. Since these volumes required a lot of memory, it was necessary to reduce the original sizes of the volumes (see Tab.4.1) and run this series of tests on a more powerful computer. A machine with a memory capacity of 16 GB finally enabled the MDL procedure to be run.

| volume no. | original size | modified size |
|---|---|---|
| 1 | $384 \times 384 \times 303$ | $76 \times 185 \times 103$ |
| 2 | $384 \times 384 \times 309$ | $76 \times 185 \times 98$ |
| 3 | $384 \times 384 \times 296$ | $76 \times 185 \times 96$ |
| 4 | $384 \times 384 \times 387$ | $76 \times 192 \times 113$ |
| 5 | $384 \times 384 \times 340$ | $76 \times 185 \times 108$ |
| 6 | $384 \times 384 \times 326$ | $76 \times 185 \times 117$ |

Table 4.1: The 6 CT hand volumes and their original as well as modified size.

## 4.2 Results

For each dataset a search over the parameter space was conducted to achieve the best outcome. In addition, several tests were made to find the best interest point detector for each of the datasets.

### 4.2.1 2D synthetic images

| Iteration | mean | median |
|---|---|---|
| (init) 0 | 5.09 | 4.96 |
| 200 | 2.44 | 0.05 |

Table 4.2: The calculated mean and median landmark error for the iteration and the optimization for 4 manually selected landmarks on the synthetic 2D dataset.

On 2D synthetic images interest points retrieved from the result of the Canny edge detector works best. The MDL results for a selection of 4 interest

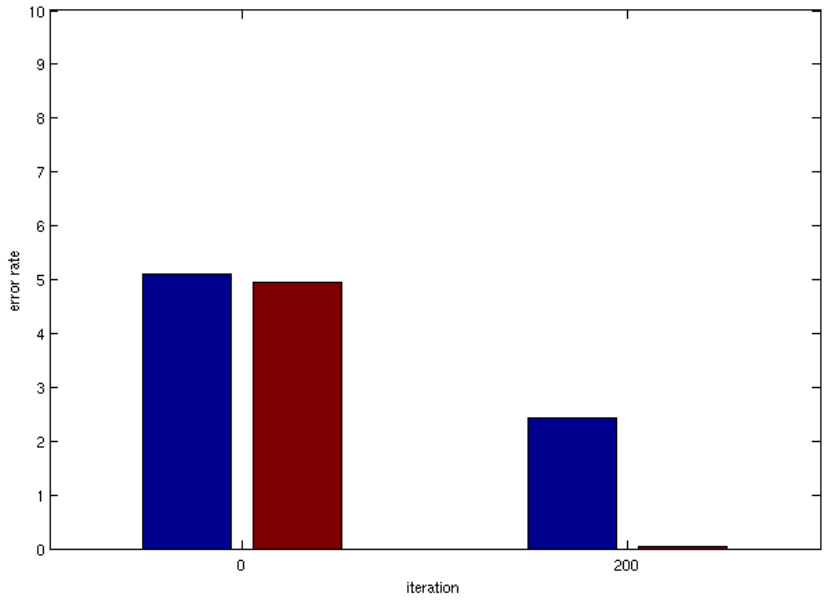



Figure 4.1: The error rates for the MDL procedure on the synthetic 2D data at the initialization (iteration 0) and after the optimization (iteration 200). The blue bars represent the mean landmark error, whereas the red bars represent the median landmark error.

points (on one corner each) after only 200 iterations can be seen in Tab. 4.2 and is also illustrated in Fig. 4.1. The initialization is visualized in Fig. 6.1 and a plot of the MDL result can be found in Fig. 6.2). The four selected corner points are found in all images except from 4 images in the last row.

Correctly retrieving the four corner points was a challenge because in the initial phase of the project, the parameters for the optimization progress were not set accordingly to the shape and texture weight. After a series of experiments, the best results could be achieved when setting the texture weight to 10 and the shape weight to 1 because then the features were put more emphasis on. That is especially important when dealing with corners!

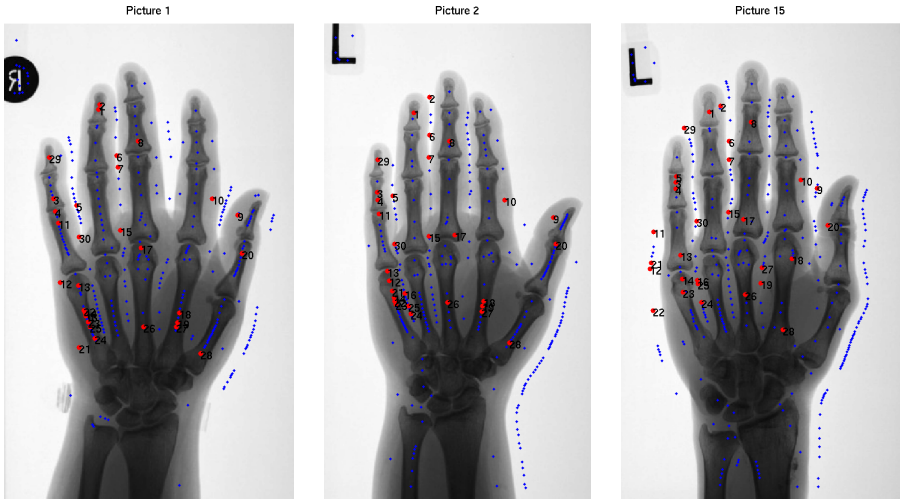
### 4.2.2 Hand X-ray images

When dealing with real 2D images (like X-ray images of human hands), symmetry interest points are the best choice to produce good results because they preserve the structure of the objects in the images. Besides, the selection of interest points is also important because in this case the MDL approach works best for a selection of 20 to 30 points equally distributed over each image.

Two experiments were conducted to demonstrate the differences within the MDL optimization process. Both of them used a set of 20 X-ray hand images with a selection of 30 landmarks and stopped after 10,000 iterations.

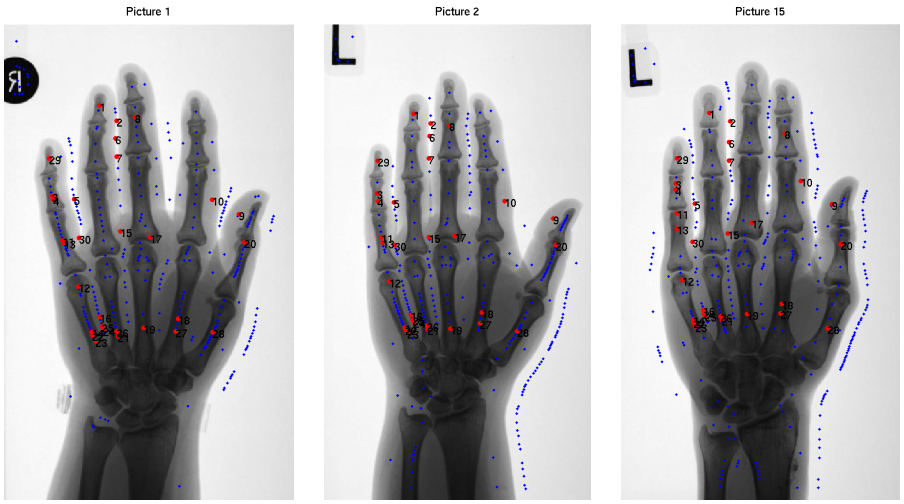
**Random selection of initial landmarks** In the first experiment the landmarks with which the MDL is initialized were chosen randomly. This initialization can be seen in Fig. 4.2 (a). After the initialization step the mean landmark error was 22.90 and the median landmark error was 19.821 (the visualization of the initialization step can be seen in Fig. 4.2). Then the optimization method was started: The corresponding error rates can be seen in Tab. 4.3 and are also visualized in Fig.4.6. The initialization step took 8 seconds, optimization procedure for 10,000 steps took 286 seconds. The visualization of three chosen images can be found in Fig. 4.3.

**Manual selection of initial landmarks** In the other experiment the initial landmarks were selected manually. This guaranteed a better distribution of landmarks over the whole image (see Fig. 4.4 (a)). After the manual selection, the optimization was performed as with the randomly selected landmarks. A visualization of 3 images after the initialization step can be seen in Fig. 4.4 with a mean landmark error of 22.80 and a median landmark error




(a) (b) (c)

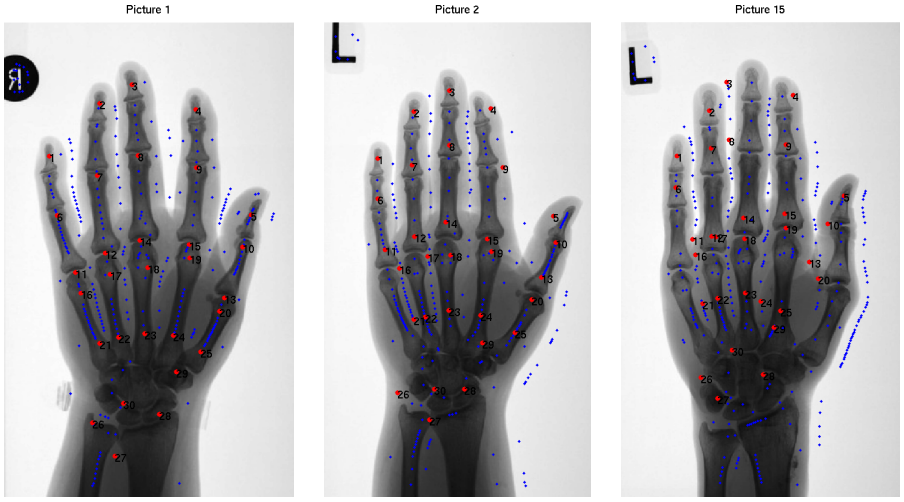
Figure 4.2: The initialization of the MDL dataset for a random selection of landmarks. Images 1 (a), 2 (b) and 15 (c) are shown. Note how the landmarks cover different sections of the hand X-ray, especially the four landmarks in image (c) outside of the hand object.




(a) (b) (c)

Figure 4.3: The visualization of the (initially randomly selected) corresponding landmarks after 10,000 iterations of the MDL optimization. Images 1 (a), 2 (b) and 15 (c) are shown. Note the correspondences between the points of the three images to see the improvement.

of 20.83. The initialization procedure took 277 seconds, the optimization for 10,000 steps procedure took 373 seconds.




(a) (b) (c)

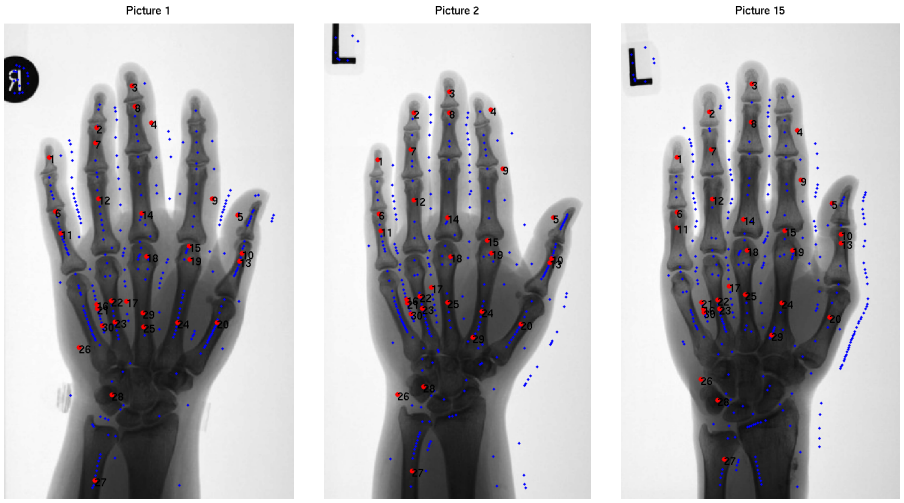
Figure 4.4: The initialization of the MDL dataset for a manual selection of landmarks. The landmarks were selected in image 1. Images 1 (a), 2 (b) and 15 (c) are shown. Note how the landmarks cover different sections of the hand X-ray, especially the landmark that is supposed to be on the middle finger in image (c)

The MDL result after 10,000 steps of the optimization is shown in Fig. 4.5. The error rates can be found in Tab. 4.3 or Fig. 4.6. As you can see, both error rates for manually and randomly selected landmarks in the initialization step of the MDL procedure are decreasing and close to each other.

The SAMs are then tested with both of the MDL results. First the MDL result for the manual selection of landmarks is used to build the SAMs. During the subsequent leave-one-out-test the error rates are computed for each image by mapping the image that is currently regarded as target image to the others. The resulting mean and median landmark error for this match is then computed and stored for this leave-one-out-test. The error rates are illustrated in Fig. 4.7 (a). The visualization of the SAM result can be seen in Fig. 4.9.

The other experiment evaluated the SAMs' success when using the MDL result of randomly selected landmarks. The error rates for the leave-one-out-tests are given in Fig. 4.7 (b) and are computed the same way as above. A visualization of the SAM result is shown in Fig. 4.8.

As you can see in Fig. 4.7, the SAMs that are build out of the MDL



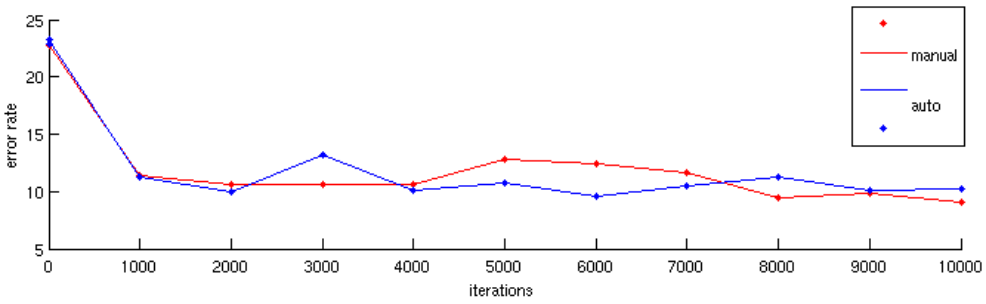

(a) (b) (c)

Figure 4.5: The visualization of the (initially manually selected) corresponding landmarks after 10,000 iterations of the MDL optimization. Images 1 (a), 2 (b) and 15 (c) are shown. Note the correspondences between the points of the three images to see the improvement.

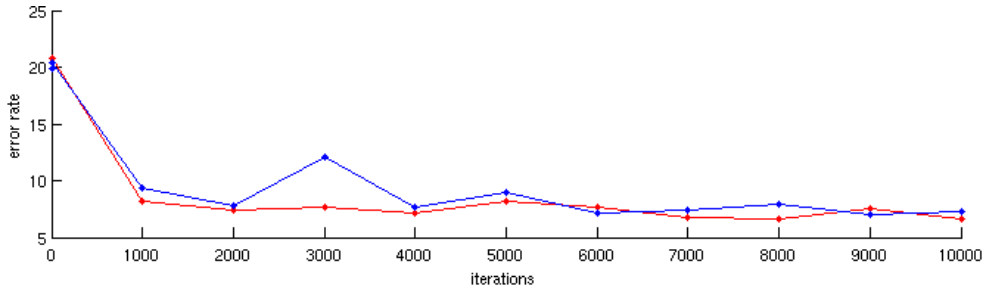
| Iteration | manual selection | | auto selection | |
|---|---|---|---|---|
| | mean | median | mean | median |
| (init) 0 | 22.80 | 20.83 | 22.90 | 19.82 |
| 1 | 22.78 | 20.74 | 23.24 | 20.44 |
| 1000 | 11.40 | 8.17 | 11.28 | 9.37 |
| 2000 | 10.62 | 7.45 | 10.03 | 7.77 |
| 3000 | 10.67 | 7.66 | 13.19 | 12.03 |
| 4000 | 10.67 | 7.15 | 10.06 | 7.68 |
| 5000 | 12.80 | 8.20 | 10.77 | 9.00 |
| 6000 | 12.46 | 7.61 | 9.63 | 7.15 |
| 7000 | 11.70 | 6.72 | 10.45 | 7.38 |
| 8000 | 9.49 | 7.58 | 11.28 | 7.88 |
| 9000 | 9.83 | 7.52 | 10.09 | 7.00 |
| 10000 | 9.01 | 6.60 | 10.18 | 7.31 |

Table 4.3: Hand X-ray images: the calculated mean and median landmark error during the iteration steps of the optimization for 30 randomly and manually selected landmarks.

result with manual selected landmarks showed a better results with only two outliers (picture 5 and 20). The other experiment with the MDL result of

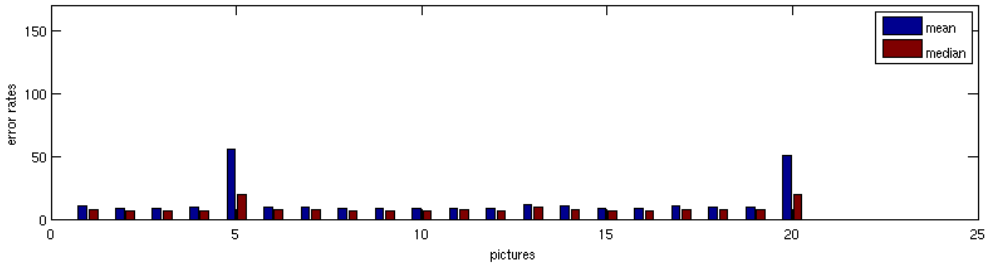



(a)

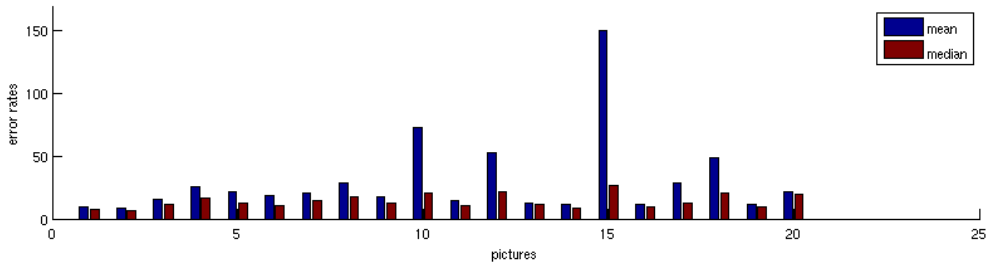



(b)

Figure 4.6: The error rates for the MDL procedure. Iteration 0 symbolizes the errors after the initialization. (a) shows the calculated mean landmark error for the manual (red line) and random (blue line) selection of landmarks. (b) illustrates the median landmark error for the manual (red line) and random (blue line) selection of landmarks.



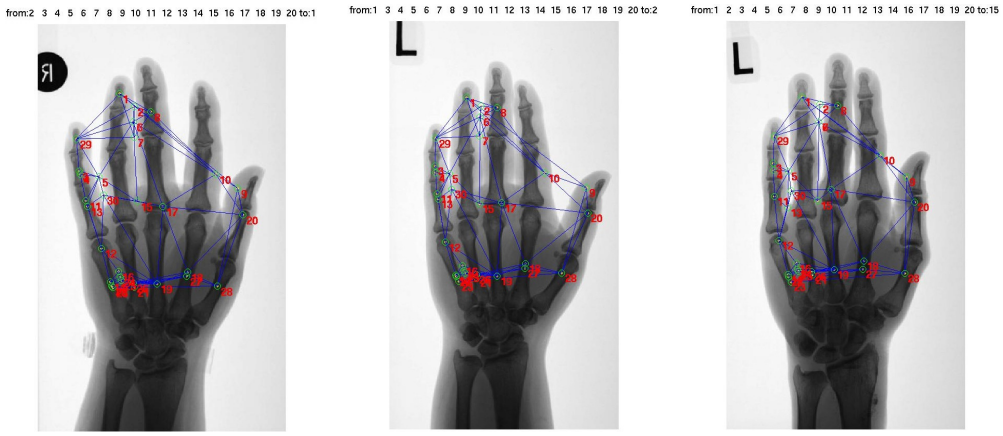

(a)




(b)

Figure 4.7: The calculated error rates while mapping the knowledge retrieved from the SAMs to each of the images. The blue bars represent the mean landmark error, whereas the red bars represent the median landmark error. (a) The error rates for the SAMs built out of the MDL result for manually selected landmarks and (b) for randomly selected landmarks.

randomly selected landmarks as basis for the SAMs had three outliers with a mean landmark error greater than 30 (picture 10, 12 and 18) and another enormous outlier (picture 15) which caused the mean landmark error for this leave-one-out-test to boost to 150.




(a) (b) (c)

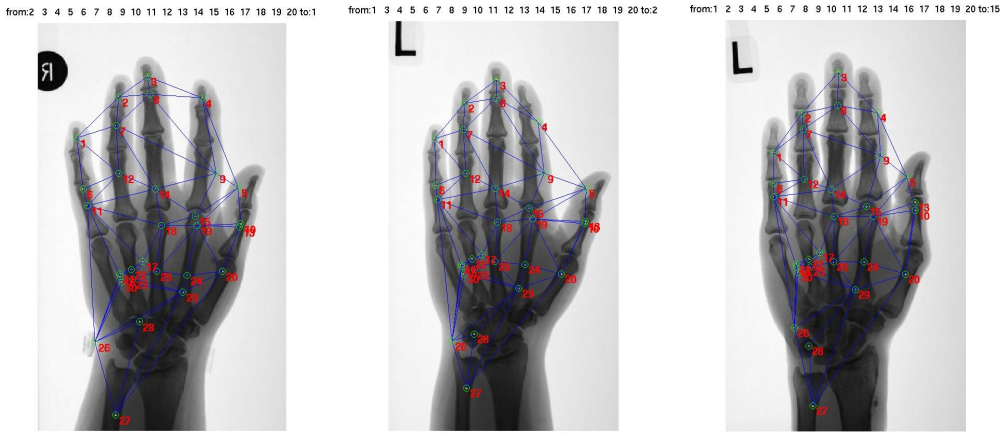
Figure 4.8: The result of three leave-one-out-tests. The knowledge computed by SAMs that are built out of the MDL result for 30 randomly selected landmarks is mapped to image 1 (a), image 2 (b) and image 15 (c). Note the high amount of correctly placed corresponding landmarks.

As you can see in Tab. 4.3, the error rates for both random and manual selections of landmarks are decreasing and close to each other. As a result, the manual selection of landmarks did not show significant improvements apart from 2 more outliers compared to the random selection.

### 4.2.3 Hand CT volumes

Final tests with the MDL procedure on the CT hand volumes showed similar results for a random selection of 25 or 50 landmarks (the total amount of symmetry interest points varied between 1650 and 1850 retrieved in each volume). Due to memory capacity a restriction to 25 landmarks was selected. A manual selection of landmarks was not tested because no ground truth data, i.e. volumes that are manually annotated by experts was available. Writing a tool to solve this problem is out of the scope of this project. Screen shots of the 6 original volumes can be found in Fig. 6.15 and Fig. 6.16.

To evaluate the MDL results volume 1 was mapped to each of the other volumes and afterwards visually inspected. It was not possible to compute error rates because no ground truth data was available.




(a) (b) (c)

Figure 4.9: The result of three leave-one-out-tests. The knowledge computed by SAMs that are built out of the MDL result for 30 manually selected landmarks is mapped to image 1 (a), image 2 (b) and image 15 (c). Note the high similarity between (a) and (b), whereas in (c) there are some more differences which results in the according error rates (see Fig. 4.7).

As you can see in Fig. 6.17 the MDL initialization is far away from an optimum result because of the holes (like in Fig. 6.17(c) or Fig. 6.17(g)). In contrast to this, the result of the MDL procedure after 10,000 iterations shows a smooth surface (e.g. see Fig. 6.18(c) or Fig. 6.18(g)) that bears a high resemblance to the original volume 1 (see Fig. 6.15(a) and Fig. 6.15(b)).

Unfortunately, several experiments with the SAMs revealed a huge problem: The 3D CT hand volumes (i.e. the number of interest points in each volume, the number of model landmarks and the corresponding number of edges between them) are too complex for the MAXSUM-solver. Even reducing the selection of landmarks to 25 per volume only yielded degenerate results. To compute the edges, adjacencies, etc. out of 5 images and map it to the sixth one took more than 24hours. In addition, the points that are computed by the SAMs collapsed to only 5 to 6 unique points although this should be prohibited by model costs. Consequently, future work will concentrate on the SAMs for real 3D data and improve this approach to achieve useful and valuable outcomes.

# Chapter 5

# Conclusion and Outlook

The MDL and SAM principles are two powerful tools in medical image processing. Having created a representative model with help of the MDL principle, Sparse MRF Appearance models can be applied to automatically locate anatomic structures in test images which means that no tedious, time-consuming and manual selection of interest points is necessary. Therefore the aim of this project was to combine both technologies to quickly achieve representative models of 2D and 3D images and retrieve the same structures in new and unseen images based on the a priori knowledge.

Evaluation of experiments with three different kinds of data (two 2D and a 3D dataset) showed excellent results of the MDL principle with low and thus acceptable error rates. Using the MDL results for SAMs also presented successful results for 2D datasets.

However, further research will focus on 3D volumes because the SAMs did not deliver useful results. This work was the first approach to apply SAMs to real 3D datasets. It will also be necessary to develop an evaluation framework for real 3D volumes to further improve the hitherto existing system.

# Chapter 6

# Appendix

This Chapter includes the complete results of all experiments. Descriptions are given below each image.


Figure 6.1: 20 synthetic images with 4 manually selected landmarks (as a selection of all points computed by the Canny edge detector) after the initialization (before the optimization starts). The 4 landmarks were selected in image number 1 as corner points.


Figure 6.2: The MDL result of 20 synthetic images. The corresponding landmarks are marked as red dots in each of the 20 synthetic images. The optimization stopped after 200 iterations. The according error rates can be found in Tab. 4.2

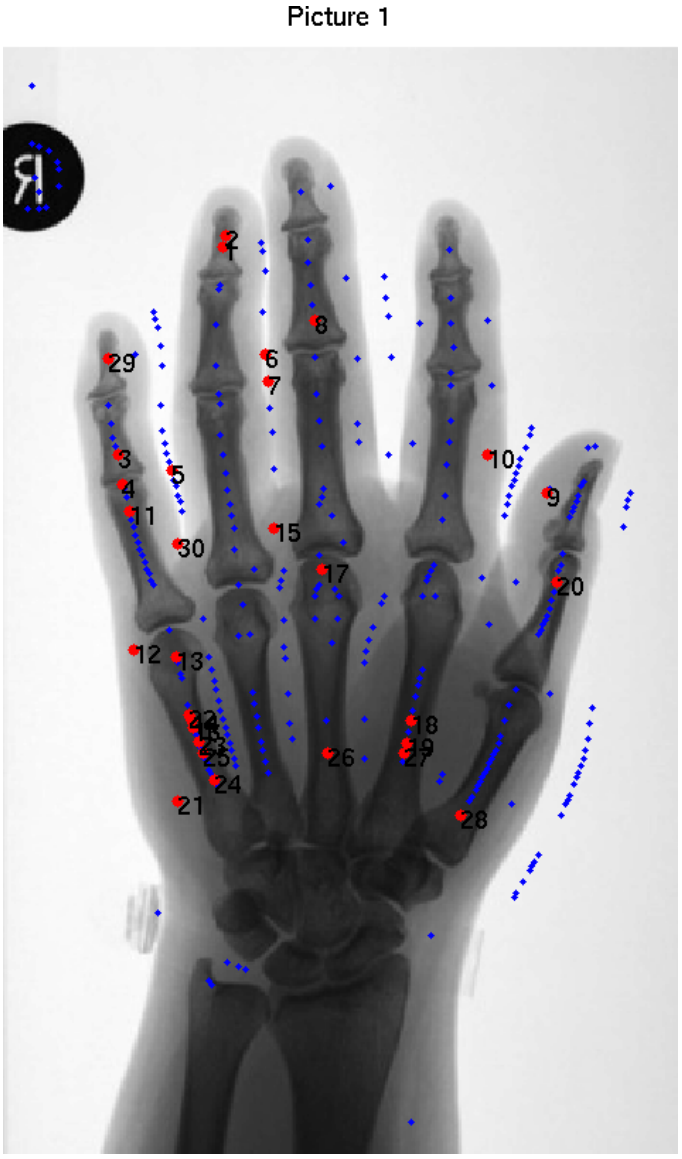



Figure 6.3: The initial (random) selection of 30 landmarks in image 1.The red dots symbolize the selected ones, the blue dots represent all retrieved symmetry interest points.

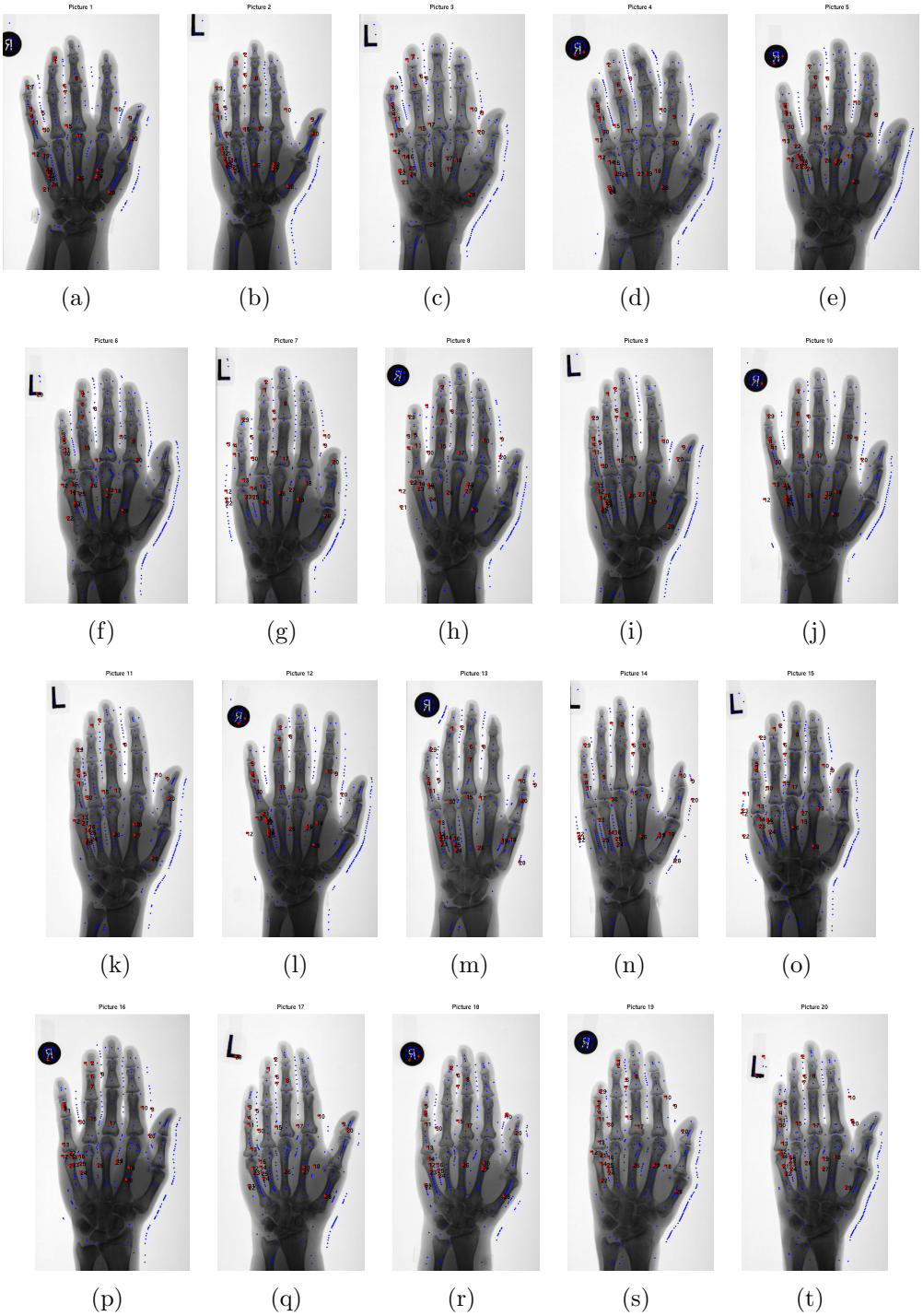



Figure 6.4: MDL-initialization of 20 X-ray images of human hands. The red dots symbolize the selected landmarks.

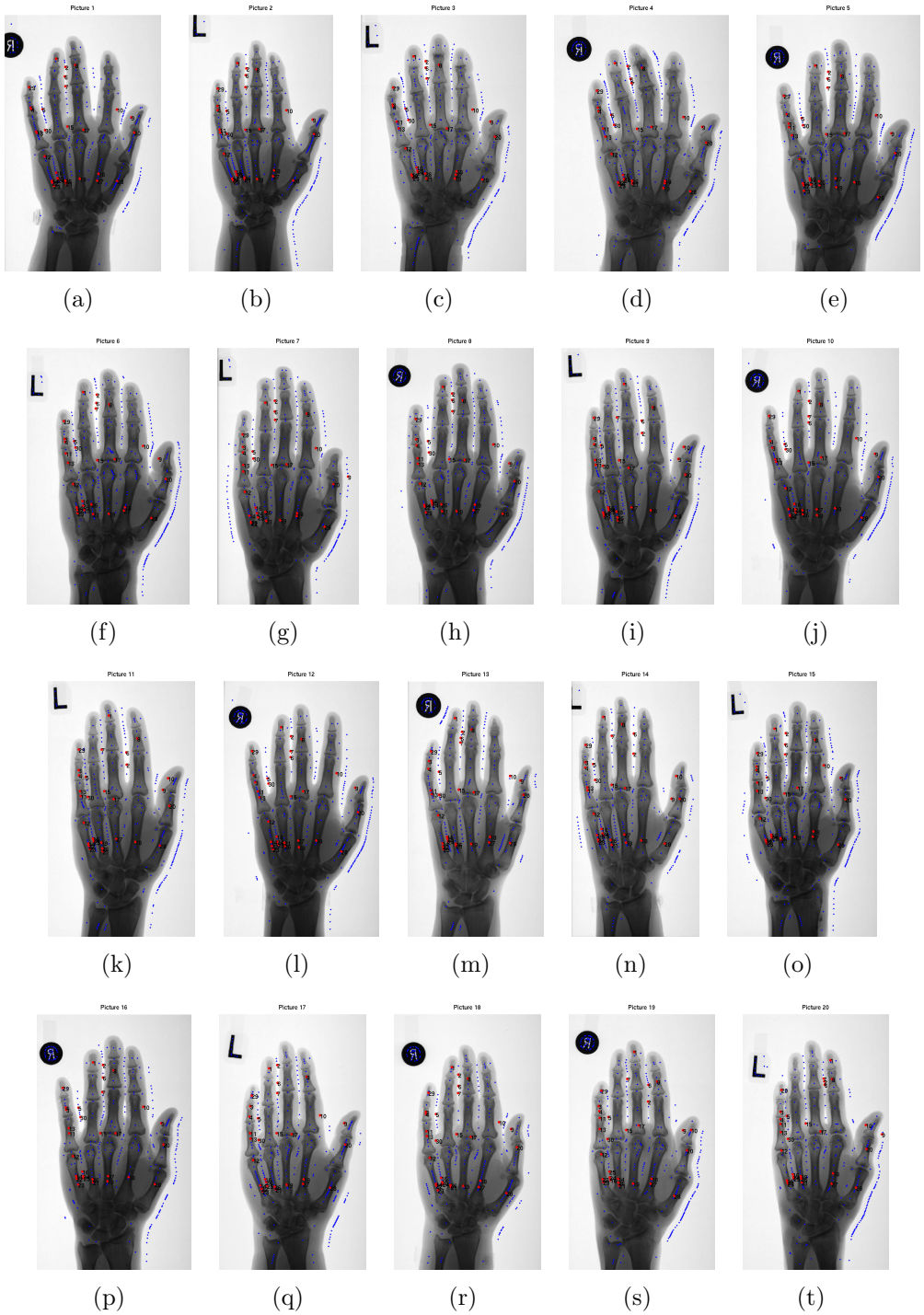
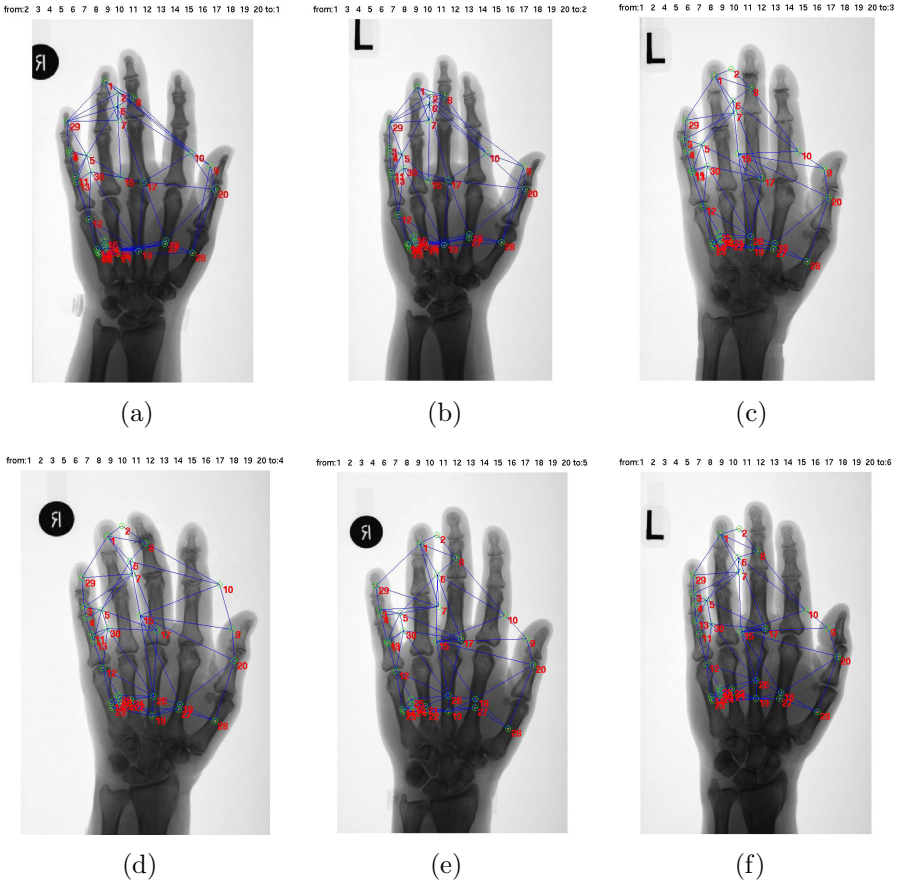


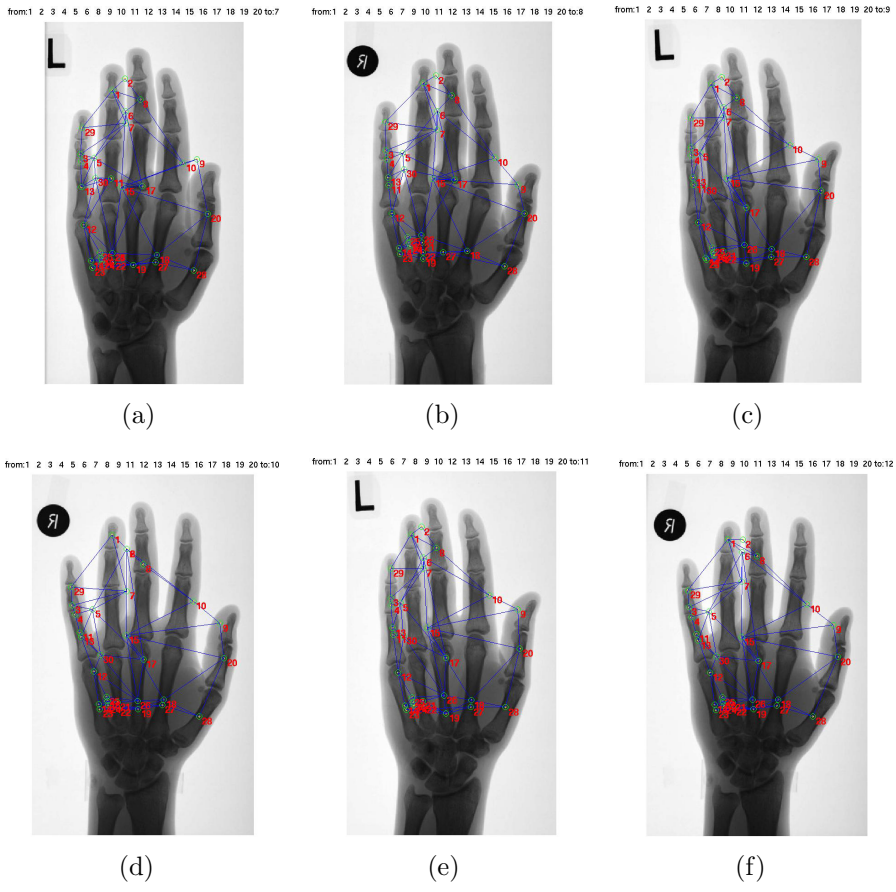
Figure 6.5: MDL result of 20 X-ray images of human hands after 10,000 iterations for a random selection of 30 landmarks. The red dots symbolize the corresponding landmarks, the blue dots represent all retrieved symmetry interest points.



(a) (b) (c)

(d) (e) (f)

Figure 6.6: SAM result of X-ray images no. 1-6 of human hands with the MDL model as starting point. The red dots symbolize the corresponding landmarks.



(a) (b) (c)

(d) (e) (f)

Figure 6.7: SAM result of X-ray images no. 7-12 of human hands with the MDL model as starting point. The red dots symbolize the corresponding landmarks.

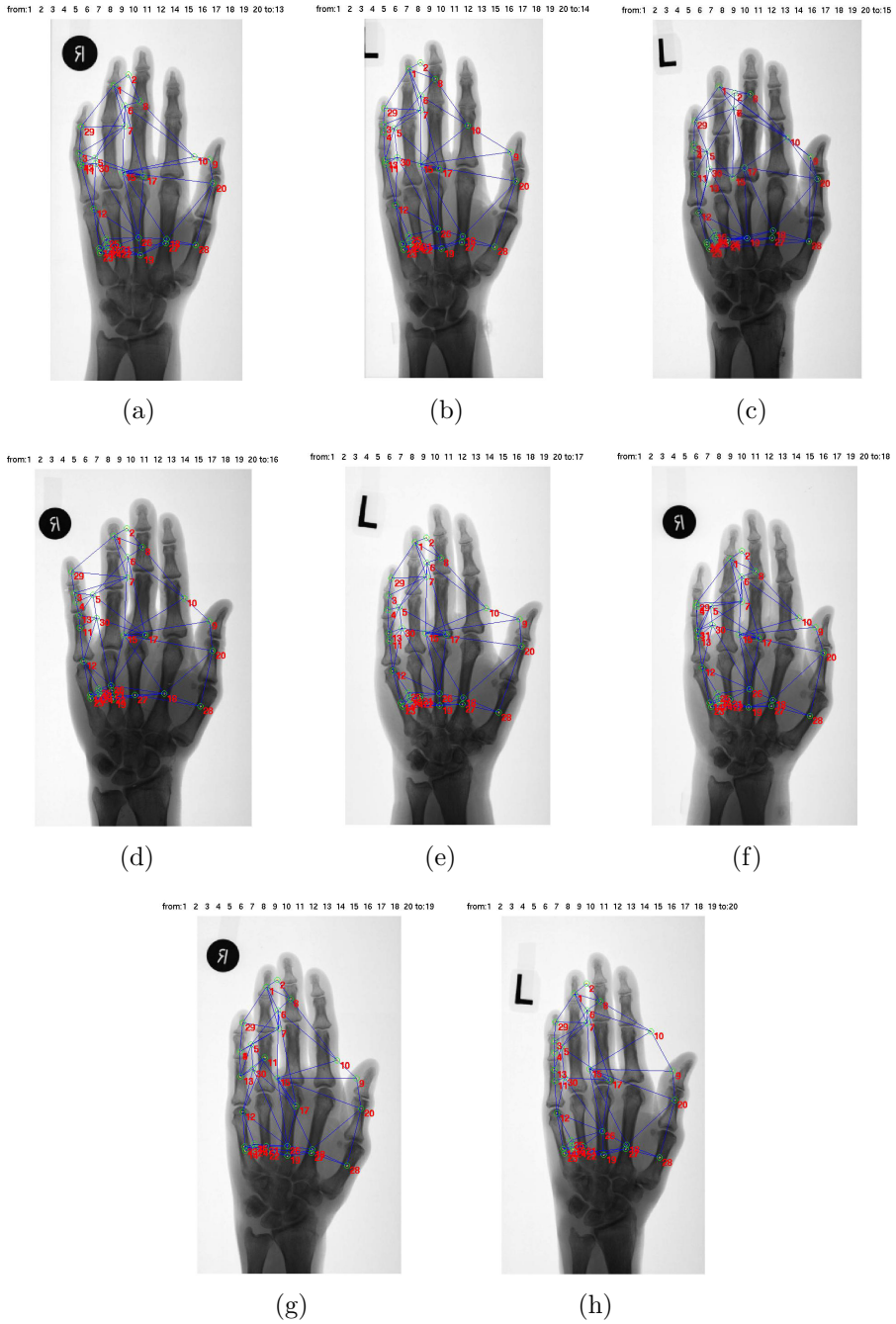


Figure 6.8: SAM result of X-ray images no. 13-20 of human hands with the MDL model as starting point. The red dots symbolize the corresponding landmarks, the blue dots represent all retrieved symmetry interest points.

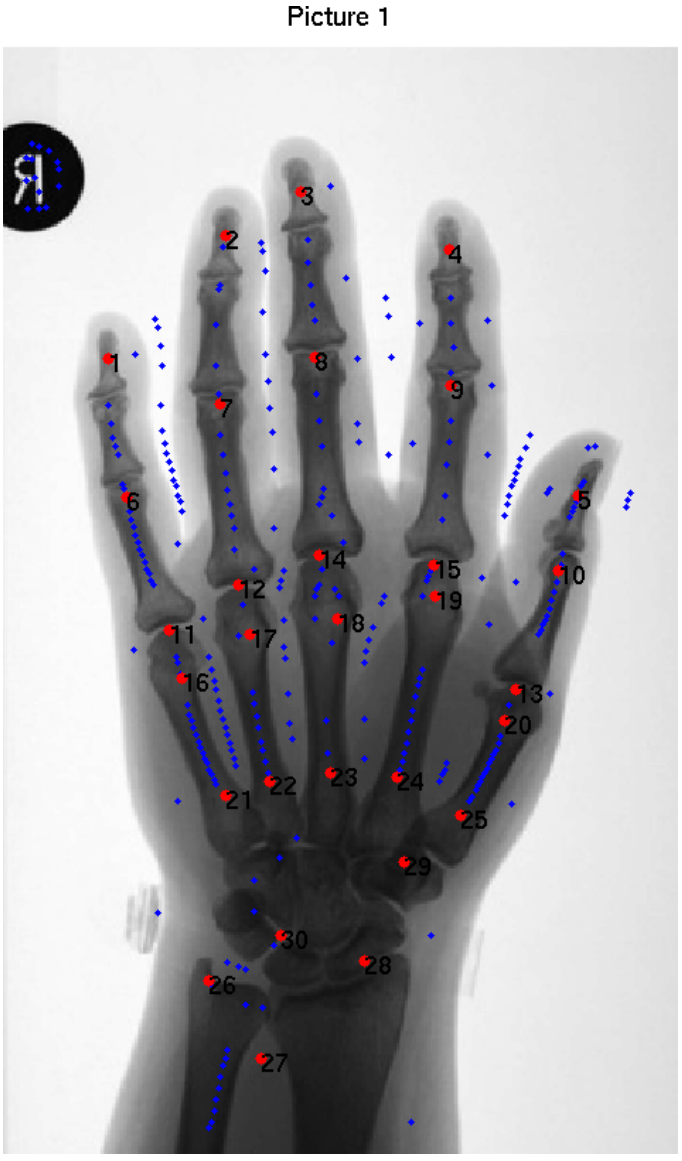



Figure 6.9: The initial (manual) selection of 30 landmarks chosen in image 1. The red dots symbolize the selected ones, the blue dots represent all retrieved symmetry interest points.

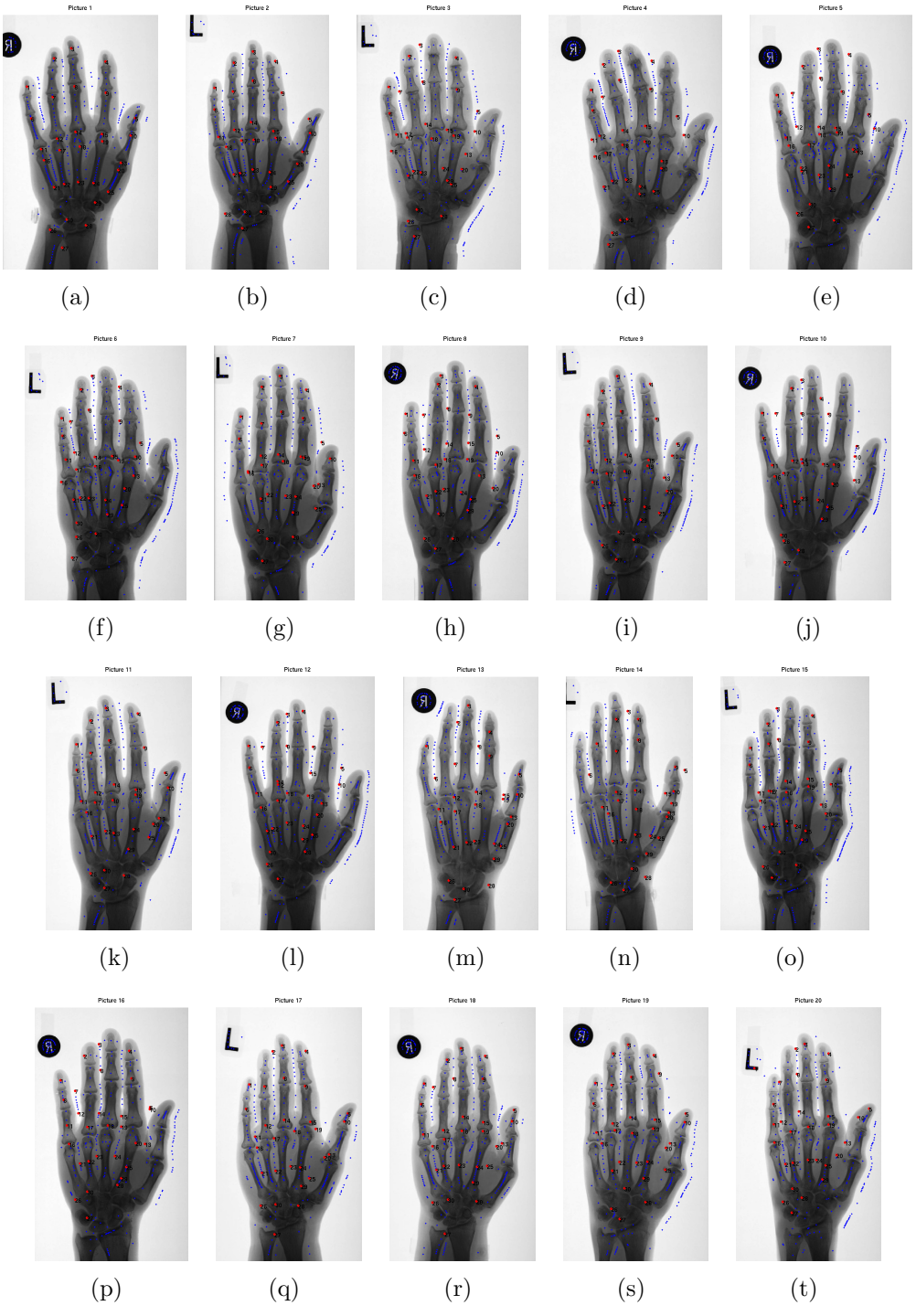



Figure 6.10: MDL-initialization of 20 X-ray images of human hands with a manual selection of 30 landmarks. The red dots symbolize the selected landmarks, the blue dots represent all retrieved symmetry interest points.

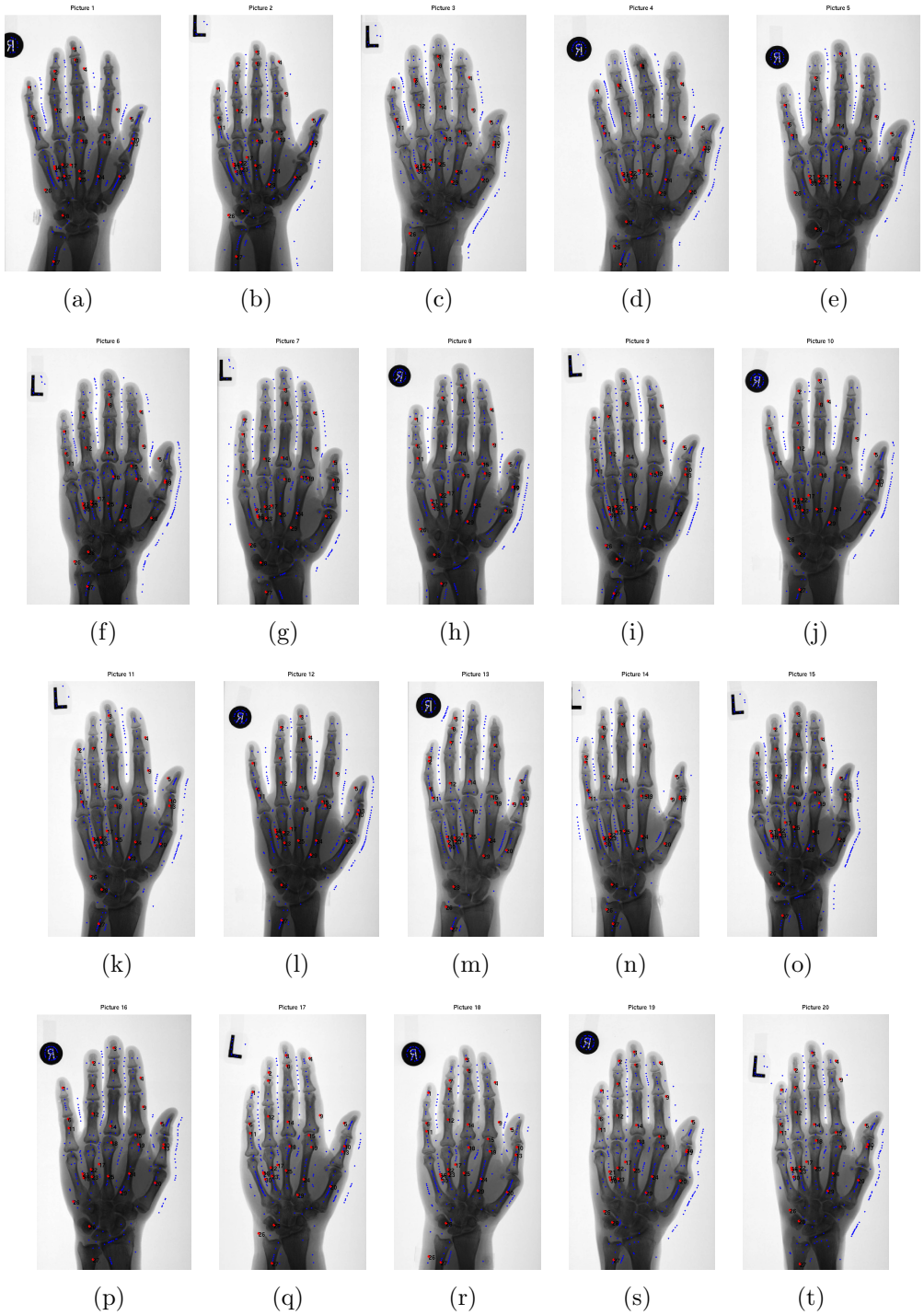
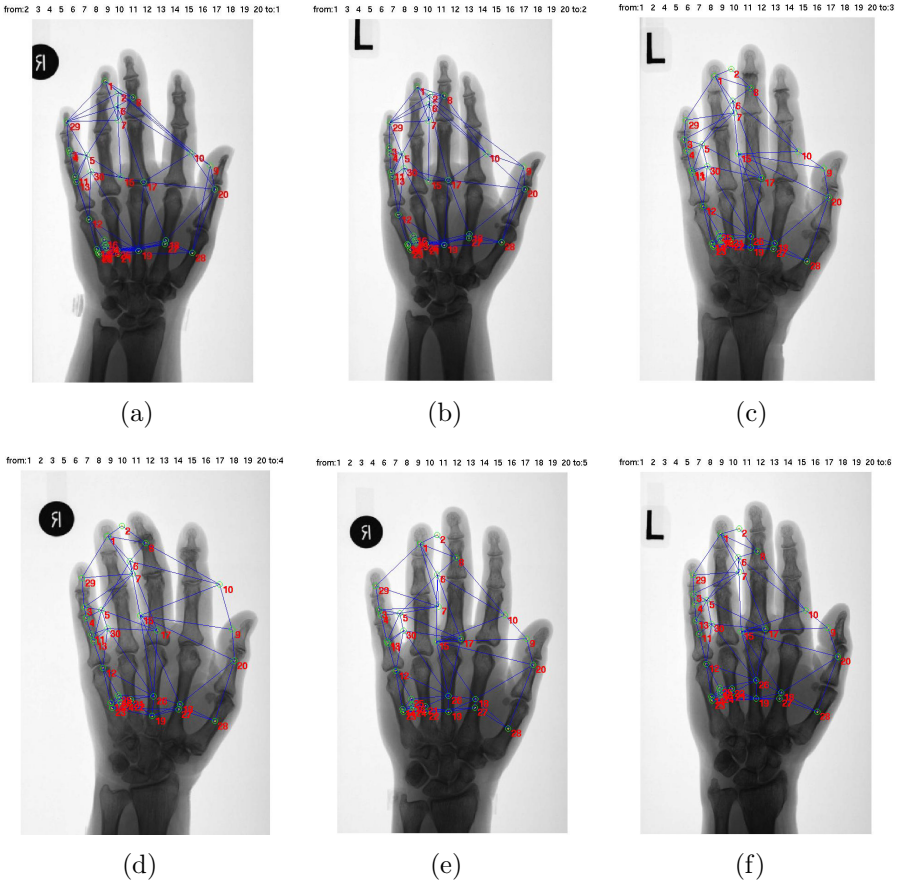


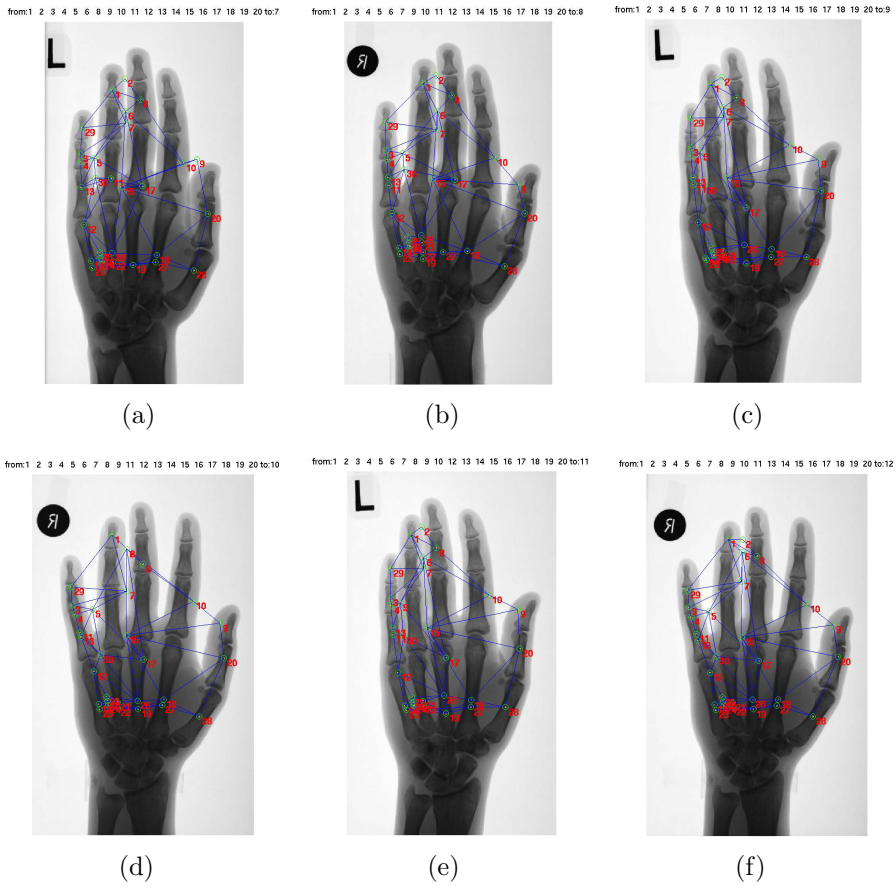
Figure 6.11: MDL result of 20 X-ray images of human hands after 10,000 iterations for a manual selection of 30 landmarks. The red dots symbolize the corresponding landmarks, the blue dots represent all retrieved symmetry interest points.



(a) (b) (c)

(d) (e) (f)

Figure 6.12: SAM result of X-ray images no. 1-6 of human hands with the MDL model as starting point. The red dots symbolize the corresponding landmarks.



(a) (b) (c)

(d) (e) (f)

Figure 6.13: SAM result of X-ray images no. 7-12 of human hands with the MDL model as starting point. The red dots symbolize the corresponding landmarks.

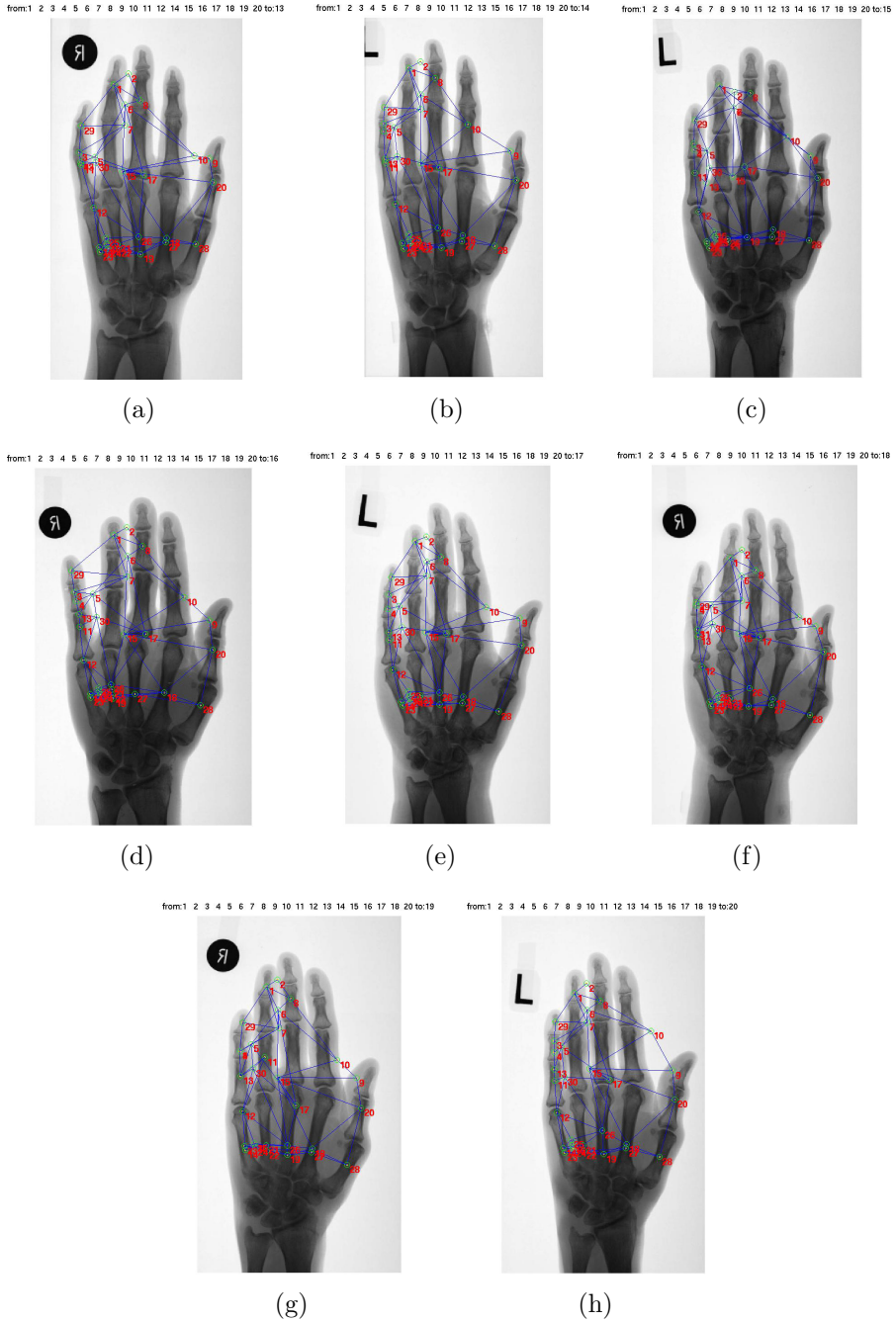


Figure 6.14: SAM result of X-ray images no. 13-20 of human hands with the MDL model as starting point. The red dots symbolize the corresponding landmarks.

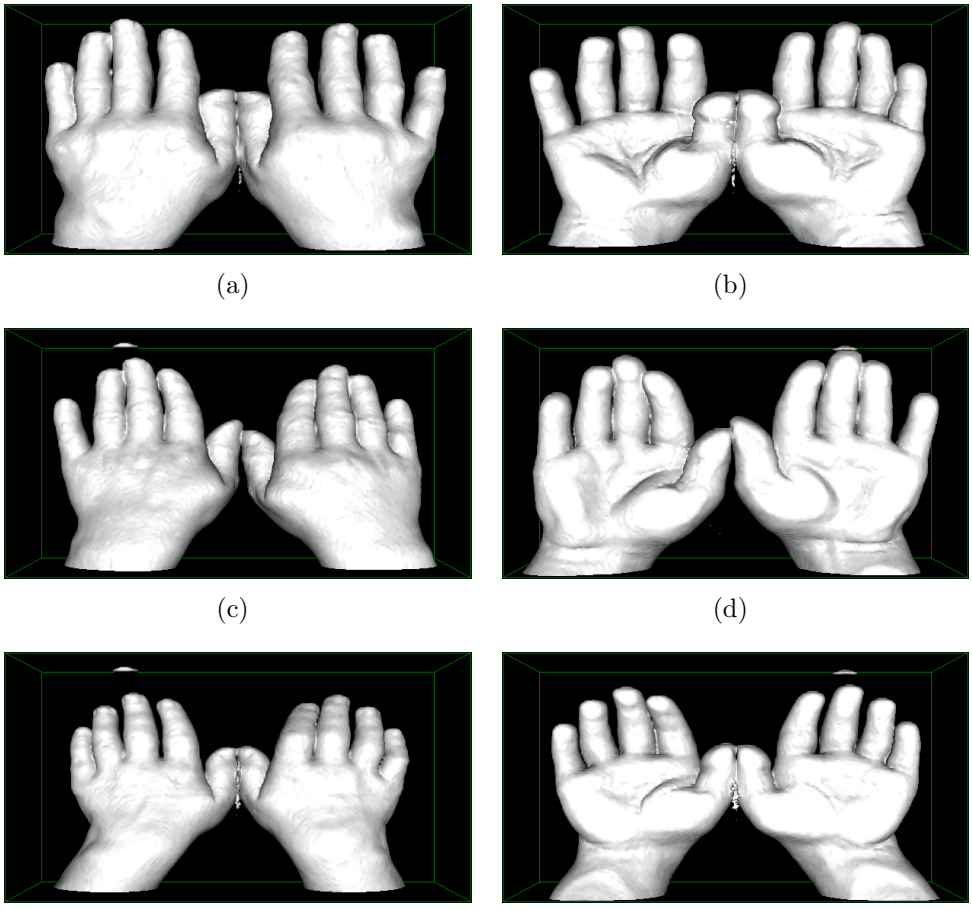
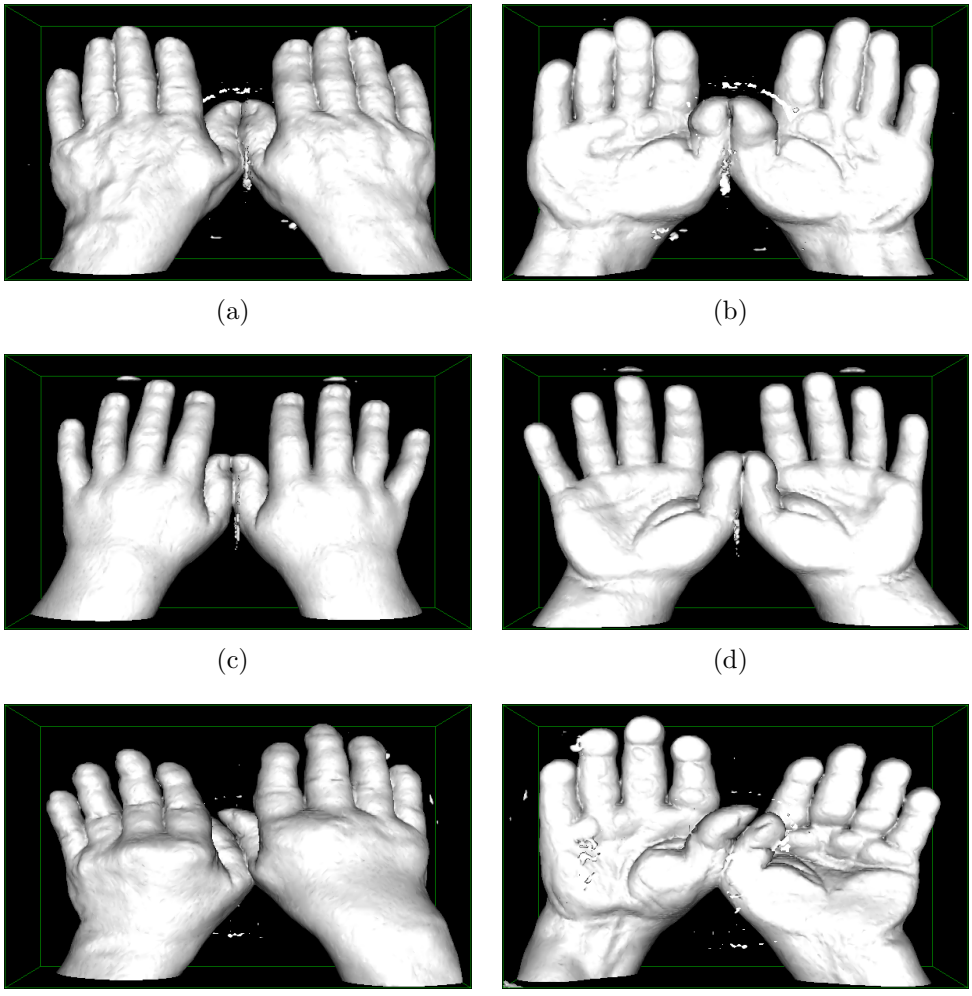


Figure 6.15: The original 3D Hand CTs number 1 to 3 with front and back view each visualized with OSIRIX [17]). a, b: volume 1. c, d: volume 2. e, f: volume 3.



(a) (b) (c) (d) (e) (f)

Figure 6.16: The original 3D Hand CTs number 4 to 6 with front and back view each. a, b: volume 1. c, d: volume 2. e, f: volume 3.

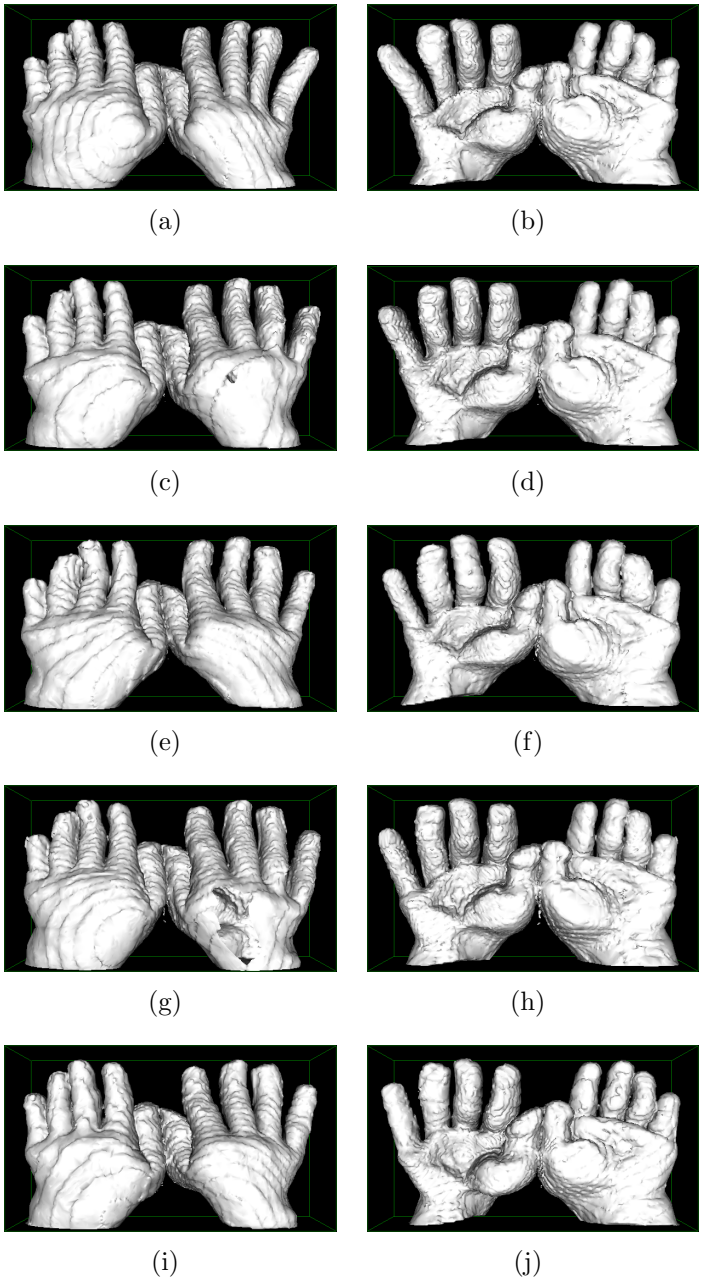



Figure 6.17: The initialization of the 3D Hand CTs before the optimization process. 25 landmarks were selected out of about 1600 interest points). Volume 1 was mapped to volume 2 (a, b), volume 3 (c, d), volume 4 (e, f), volume 5 (g, h) and volume 6 (i, j).

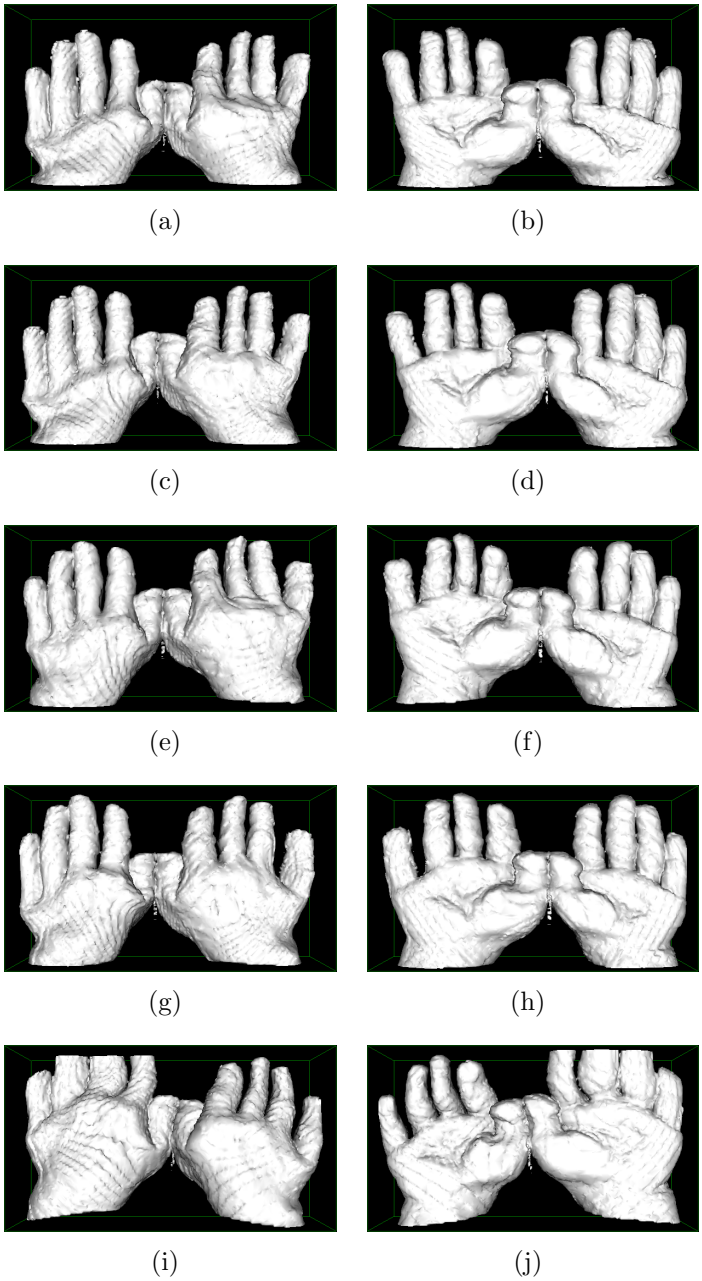



Figure 6.18: The results of the MDL procedure for the 3D Hand CTs after 10000 iterations. 25 landmarks were selected out of about 1600 interest points. Volume 1 was mapped to volume 2 (a, b), volume 3 (c, d), volume 4 (e, f), volume 5 (g, h) and volume 6 (i, j).

# Bibliography


[1] Rene Donner, Branislav Micusik, Georg Langs, and Horst Bischof. Sparse MRF Appearance Models for Fast Anatomical Structure Localisation. *BMVC*, 2007.

[2] R Donner, B Micusik, G Langs, and L Szumilas. Object Localization Based on Markov Random Fields and Symmetry Interest Points. *LECTURE NOTES IN COMPUTER SCIENCE*, Jan 2007.

[3] Tim F. Cootes and Chris J. Taylor. Active Shape Models - Smart Snakes. In *Proceedings of the British Machine Vision Conference*, pages 266–275, September 1992.

[4] Tim F. Cootes, Gareth J. Edwards, and Chris J. Taylor. Active Appearance Models. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, volume 23, pages 681–685, June 2001.

[5] D. Freedman and T. Zhang. Interactive graph cut based segmentation with shape priors. In *IEEE CVPR*, volume 1, pages 755–762, 2005.

[6] D. M. Greig, B. T. Porteous, and A. H. Seheult. Exact Maximum A Posteriori Estimation for Binary Images. *Journal of the Royal Statistical Society*, pages 271–279, 1989.

[7] Mikkel Bille Stegmann. Active Appearance Models. Master's thesis, Technical University of Denmark, 2000.

[8] Georg Langs, Rene Donner, Philipp Peloschek, and Horst Bischof. Robust Autonomous Model Learning from 2D and 3D Data Sets. In *MICCAI PART 1*, pages 968–976, 2007.

[9] Peter Grünwald. A Tutorial Introduction to the Minimum Description Length Principle. *MIT Press*, 2005.

[10] Rhodri H. Davies, Tim F. Cootes, and Chris J. Taylor. A Minimum Description Length Approach to Statistical Shape Modelling. *IEEE Transactions on Medical Imaging*, 21(5):525–537, May 2002.

[11] Peter Grünwald. *The Minimum Description Length Principle*. The MIT Press, 2007.

[12] John Canny. A Computational Approach To Edge Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6):679–714, November 1986.

[13] Chenyang Xu and Jerry L. Prince. Gradient Vector Flow: A New External Force for Snakes. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, page 66, 1997.

[14] Jorma Rissanen. Modeling by shortest data description. *Automatica*, 14(5):465–471, September 1978.

[15] Paul Vitányi and Ming Li. Minimum Description Length Induction, Bayesianism, and Kolmogorov Complexity. *IEEE Transactions on Information Theory*, 46(2):446–464, November 2000.

[16] Rohan Baxter and Jonathan Oliver. MDL and MML: Similarities and Differences. Technical report, Department of Computer Science, Monash University, Australia, 1994.

[17] The OsiriX Foundation. OsiriX - DICOM Viewer. Website, 09/29/2008 at 4:54 pm. `http://www.osirix-viewer.com/`.