

Probabilistic Appearance Models for Segmentation and Classification

Julia Krüger

Jan Ehrhardt

Heinz Handels

{krueger, ehrhardt, handels}@imi.uni-luebeck.de

Institute of Medical Informatics, University of Lübeck, Germany

Abstract

Statistical shape and appearance models are often based on the accurate identification of one-to-one correspondences in a training data set. At the same time, the determination of these corresponding landmarks is the most challenging part of such methods. Hufnagel et al. developed an alternative method using correspondence probabilities for a statistical shape model. We propose the use of probabilistic correspondences for statistical appearance models by incorporating appearance information into the framework. A point-based representation is employed representing the image by a set of vectors assembling position and appearances. Using probabilistic correspondences between these multi-dimensional feature vectors eliminates the need for extensive preprocessing to find corresponding landmarks and reduces the dependence of the generated model on the landmark positions. Then, a maximum a-posteriori approach is used to derive a single global optimization criterion with respect to model parameters and observation dependent parameters, that directly affects shape and appearance information of the considered structures. Model generation and fitting can be expressed by optimizing the same criterion. The developed framework describes the modeling process in a concise and flexible mathematical way and allows for additional constraints as topological regularity in the modeling process. Furthermore, it eliminates the demand for costly correspondence determination. We apply the model for segmentation and landmark identification in hand X-ray images, where segmentation information is modeled as further features in the vectorial image representation. The results demonstrate the feasibility of the model to reconstruct contours and landmarks for unseen test images. Furthermore, we apply the model for tissue classification, where a model is generated for healthy brain tissue using 2D MRI slices. Applying the model to images of stroke patients the probabilistic correspondences are used to classify between healthy and pathological structures. The results demonstrate the ability of the probabilistic model to recognize healthy and pathological tissue automatically.

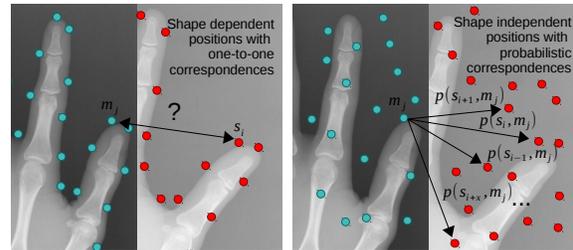


Figure 1. In classical SSMs or AAMs the sampling point positions are chosen to represent the shape and/or distinctive landmarks (left) and correct one-to-one correspondences are needed. In our approach (right) the sampling points are distributed independently of the object shape and probabilistic correspondences $p(s_i | m_j)$ between a point $m_j \in M$ and all the other sampling points $s_i \in S$ are computed.

1. Introduction

Statistical models play an important role in image segmentation and classification [10]. Especially in medical image analysis Statistical Models address the heterogeneous nature of anatomical structures with high variability in shape and appearances. Cootes *et al.* [3] introduced a Statistical Shape Model (SSM), where the shapes of objects are represented by a set of landmarks describing statistical shape properties. Classical Shape Models are generated using a principal component analysis (PCA). To determine the mean shape and possible shape variations of the considered structure included in a training data set, the corresponding landmarks that represent the same anatomical location have to be known. A manual identification of these points is not practical due to their large number. Therefore, automatic approaches were proposed in the literature for the determination of one-to-one corresponding landmarks [1, 10, 17, 20]. Furthermore, it is debatable whether the exact same anatomical landmark exists in all images at well defined positions due to anatomical differences or different acquisition techniques for medical images.

Point-based registration approaches like the Iterative Closest Point (ICP) algorithm encounter the same problem of non-identical or non-existing point-pairs. Therefore, algorithms have been developed using probabilistic instead

of one-to-one correspondences. These probability-based approaches include the Expectation Maximization-Iterative Closest Point (EM-ICP) algorithm by Granger *et al.* [9], the Softassign [8], or the Coherent Point Drift algorithm [18] and other non-rigid approaches [2, 13].

Hufnagel *et al.* [12, 11] proposed the use of correspondence probabilities when building an SSM to model the uncertainty of the point locations and to reduce the costs for determining the “correct” landmarks. A maximum-a-posteriori (MAP) framework was developed estimating the model parameters (mean shape and variation modes) by minimization of a single global criterion. Active Appearance Models (AAM) were developed as an extension of SSMs, modeling also the appearance information of an object [4]. However, AAMs equally presuppose the existence of one-to-one correspondences and are generated in three steps: (1) identification of corresponding landmarks and generation of an SSM, (2) warping all images to the mean shape and sampling image intensities to build the appearance model, and (3) combining both models. In the model adaption process to an unseen test image, another warping step is necessary for shape-normalization before computing intensity differences [4].

The idea of using appearance models for segmentation and classification of previously trained objects is similar to atlas-based approaches [19]. But the problem to be solved with most atlas-based (therefore registration-based) approaches is the presupposed existence of all classes/segmentation regions. The challenge of registering occluded images with un-occluded images is not fully solved yet. Therefore, using a model without presupposed correspondences allows for more robustness and during the adaption/registration step not only positions but also appearances can be fitted to the new unseen image.

In this work we propose a new framework for appearance models using probabilistic correspondences, where sampling point positions and appearance features are modeled at the same time using a single global optimization criterion. We extended the method of Hufnagel *et al.* [12, 11] by appearance information. For this purpose, we use a point-based representation of images combining sampling point positions and appearance information. Then, a MAP approach is employed to derive a single global optimization criterion with respect to model parameters and observation dependent parameters, that affect shape and appearance information of the considered structures. Model generation and model fitting can be expressed by optimizing the same criterion. In contrast to classical AAMs, the preprocessing for identifying one-to-one correspondences and the warping-step to eliminate shape differences is no longer required. Furthermore, we extend the derived criterion by a regularization term which penalizes implausible topological changes.

Previous work on sparse appearance models treat shape and appearance parameters separately and combine them afterwards (like AAM) or are based on complex local descriptors [5, 6]. The main idea of our point-based appearance model is that position and appearance features are treated the same way and optimized equally at the same time. Furthermore, the approach does not depend on particular appearance features at interest image points.

We applied the model for segmentation of the hand contour as well as the detection of landmarks describing finger joints in 2D hand X-ray images. Segmentation information (labels/landmarks) is modeled as further features in the vectorial image representation. The results demonstrate the feasibility of the model to reconstruct contours as well as landmarks for unseen test images. Furthermore, we applied the model for classification, where a model is generated for healthy brain tissue using 2D MRI slices. Adapting the model to pathological images containing stroke tissue the probabilistic correspondences are used to classify between healthy and pathological structures demonstrating the ability of the probabilistic model to distinguish between healthy and pathological tissue.

Our goal is to introduce a flexible approach based on a closed-form mathematical formulation, that is not depending on correspondences and is not composed of multiple steps ((1) correspondence determination, (2) building SSM, (3) warping, (4) building AAM) but on the minimization of a single global criterion.

2. Methods

2.1. Sparse Image Representation

We generate a point-based representation of an image $I : \Omega \rightarrow \mathbb{R}, \Omega \subset \mathbb{R}^{D_x}$. Under the assumption that we can compute D_f appearance features for each image position, image I is represented by N_s samples at positions $\mathbf{x}_i = (x_{i,1}, \dots, x_{i,D_x}) \in \Omega$, with appearance features $\mathbf{f}_i = (f_{i,1}, \dots, f_{i,D_f})$ for each sample i . The positions and features are now assembled in N_s vectors $\mathbf{s}_i = (\mathbf{x}_i, \mathbf{f}_i)$ of dimension $D = D_x + D_f$: $\mathbf{S} = \{\mathbf{s}_i | i = 1, \dots, N_s\}$.

In the simplest case, the appearances are the image’s gray values. But to generate application-specific representations, gradient-based, Haar-like or frequency-based features such as Gabor wavelets [15] can be used.

In contrast to classical SSMs or AAMs, where the shape is explicitly modeled by landmarks representing the shape or distinctive landmarks, the idea is to choose the sampling point positions independently of any shape information (see Fig. 1), e.g. by setting the points in a grid overlying the image, or to randomize the positions. We have investigated this approach in a previous work [7]. Here, we determine the sampling points depending on appearance features (e.g. highest Gabor wavelet filter answers) under the condition

that a certain minimal distance (spacing) is maintained between all samples. This is assumed to lead to sampling points in similar anatomical regions in all of the observations and hence improves the data reconstruction from the sparse image representation.

2.1.1 Probabilistic Correspondences

In contrast to classical SSMs and AAMs, where the corresponding point pairs between the objects of the training data set need to be predefined, the determination of correspondences or rather correspondence probabilities is part of the model generation in our approach. Given two representations $\mathbf{S} = \{\mathbf{s}_i | i = 1, \dots, N_s\}$ (observation) and $\mathbf{M} = \{\mathbf{m}_j | j = 1, \dots, N_m\}$ (model), the idea is to determine the probability that sampling point \mathbf{s}_i is a noisy observation of model point \mathbf{m}_j (see Fig. 1). This can be defined using a multivariate Gaussian distribution:

$$p(\mathbf{s}_i | \mathbf{m}_j) = \frac{1}{(2\pi)^{\frac{D}{2}} |\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\mathbf{s}_i - \mathbf{m}_j)^T \Sigma^{-1} (\mathbf{s}_i - \mathbf{m}_j)\right), \quad (1)$$

where $\Sigma \in \mathbb{R}^{D \times D}$ contains the covariances of the point positions and appearance features (diagonal values, >0). It should be noted that in contrast to one-to-one correspondences, the number of sampling points N_s in \mathbf{S} does not have to match the number of sampling points N_m in \mathbf{M} .

Expectation Maximization (EM) Approach for Correspondence Determination: Given \mathbf{S} and \mathbf{M} without known correspondences and a set of parameters θ describing a transformation of \mathbf{M} , the goal is now to find the best parameters θ to match \mathbf{S} and \mathbf{M} . To solve this problem, Granger *et al.* [9] used an EM approach to perform a point-based registration. It maximizes the expectation of \mathbf{S} for a given \mathbf{M} and θ :

$$\mathbb{E}(\log p(\mathbf{S} | \mathbf{M}, \theta)) = \frac{1}{N_m} \sum_{i=1}^{N_s} \sum_{j=1}^{N_m} \mathbb{E}(\mathbf{H}_{ij}) \log p(\mathbf{s}_i | \mathbf{m}_j, \theta), \quad (2)$$

under the assumption that \mathbf{s}_i are independent of each other and with $\mathbb{E}(\mathbf{H}_{ij})$ representing the expectation that \mathbf{s}_i corresponds to \mathbf{m}_j for a given θ :

$$\mathbb{E}(\mathbf{H}_{ij}) = p(\mathbf{s}_i | \mathbf{m}_j, \theta) / \sum_{\iota=1}^{N_m} p(\mathbf{s}_i | \mathbf{m}_\iota, \theta). \quad (3)$$

The EM approach now maximizes the likelihood iteratively, where it alternates between an expectation step, in which $\mathbb{E}(\mathbf{H}_{ij})$ is estimated for the current parameters θ , and a maximization step, where the estimated likelihood is maximized with respect to θ [9].

2.2. Probabilistic Appearance Model (PAM)

The aim of the model generation is to construct a generative appearance model that optimally fits a given data

IN N observations (with N_f features); number of variation modes n

INITIALIZE $\mathbf{S}_k = \{\mathbf{s}_{k,i} = (\mathbf{x}_{k,i}, \mathbf{f}_{k,i}) | i = 1, \dots, N_s(k)\}$

DO WHILE $C(Q, \Theta)$ decreases

Optimize model parameter (with fix Q)

DO WHILE $C(Q, \Theta)$ decreases

1. estimate variances Σ between \mathbf{S}_k and \mathbf{M}_k
2. optimize $\bar{\mathbf{M}}$
3. optimize \mathbf{v}_p and $\lambda_p, p = 1, \dots, n$

Optimize observation parameter (with fix Θ)

DO WHILE $C(Q_k, \Theta)$ decreases for each k

4. optimize T_k (and update $\mathbf{M}_k, \mathbb{E}(\mathbf{H}_{ij})_k$)
5. optimize ω_k (and update $\mathbf{M}_k, \mathbb{E}(\mathbf{H}_{ij})_k$)

OUT \mathbf{M}_k

OUT $\Theta = \{\bar{\mathbf{M}}, \mathbf{v}_p, \lambda_p, n, \Sigma | p = 1, \dots, n\}$

Figure 2. Model generation: First model parameters Θ are optimized until convergence (step 1.-3.), then the observation parameters are optimized with fixed Θ (step 4.-5.). The model adaption process consists only of the underlaid steps 4-5 with output \mathbf{M}_k .

set of N images with associated image representations $\mathbf{S}_k, k = 1, \dots, N$. As for AAMs, our PAM consists of a mean model $\bar{\mathbf{M}}$ and a set of n variation modes \mathbf{v}_p (with standard deviations λ_p) reflecting the variation in the training data \mathbf{S}_k . The observations \mathbf{S}_k are interpreted as randomly generated by the (unknown) model. Now we can divide the parameters of our approach into model parameters $\Theta = \{\bar{\mathbf{M}}, \mathbf{v}_p, \lambda_p, n, \Sigma | p = 1, \dots, n\}$ and observation dependent parameters $Q_k = \{T_k, \omega_k\}, k = 1, \dots, N$, where $\Sigma \in \mathbb{R}^{D \times D}$ is the covariance matrix (diagonal values, >0) to regulate the correspondence between \mathbf{s}_i and \mathbf{m}_j , ω_k are the observation dependent weights of the variation modes and T_k is an affine transformation to match \mathbf{S}_k to the model instance \mathbf{M}_k , where $T_k \star \mathbf{s}_{k,i}$ only transforms the sampling positions and not the appearance information: $T_k \star \mathbf{s}_{k,i} = (T_k(\mathbf{x}_i), \mathbf{f}_i)$. Given Θ and Q_k , the model generates new instances by

$$\mathbf{M}_k = T_k^{-1} \star (\bar{\mathbf{M}} + \sum_{p=1}^n \omega_{k,p} \mathbf{v}_p), \quad \text{with} \quad (4)$$

$$\mathbf{m}_{k,j} = T_k^{-1} \star (\bar{\mathbf{m}}_j + \sum_{p=1}^n \omega_{k,p} \mathbf{v}_{p,j}). \quad (5)$$

2.2.1 Maximum A-Posteriori Approach for Parameter Optimization

To find the optimal parameters for Θ and Q_k , the idea is to optimize the model together with the observation dependent weights and the correspondences between observations

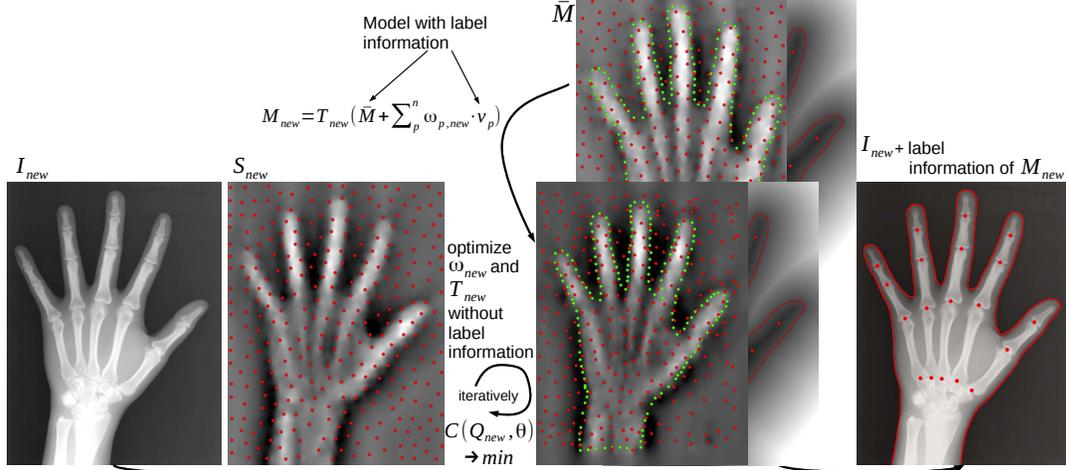


Figure 3. Model Adaption Process. Note that the observation \mathbf{S}_{new} does neither include the segmentation features nor the sampling points on the hand border (green dots) but the corresponding model instance \mathbf{M}_{new} does. Therefore, the wanted segmentation information on the right (red contour and joint landmarks) is reconstructed with additional points from the model. Furthermore, it should be noted, that the model adaption is performed by minimizing the differences between the sparse representations \mathbf{S}_{new} and \mathbf{M}_{new} and not I_{new} and \mathbf{M}_{new} .

\mathbf{S}_k and associated model instances \mathbf{M}_k . Therefore, the a-posteriori probability that all \mathbf{S}_k can be generated by Θ with Q_k is maximized: $\sum_{k=1}^N p(Q_k, \Theta | \mathbf{S}_k) \rightarrow \max$. Using Bayes' theorem we minimize:

$$\begin{aligned} C(Q, \Theta) &= -\sum_{k=1}^N \log p(Q_k, \Theta | \mathbf{S}_k) \\ &= -\sum_{k=1}^N \log \left(\frac{p(\mathbf{S}_k | Q_k, \Theta) p(Q_k | \Theta) p(\Theta)}{p(\mathbf{S}_k)} \right) \end{aligned} \quad (6)$$

Each partial probability can be considered and defined separately:

Prior probability of observations $p(\mathbf{S}_k)$: $p(\mathbf{S}_k)$ is assumed to be constant.

Likelihood $p(\mathbf{S}_k | Q_k, \Theta)$: The probability of the model instances generated by the model depends on the weights ω_k and transformation T_k of \mathbf{S}_k :

$$\begin{aligned} p(\mathbf{S}_k | Q_k, \Theta) &= p(\mathbf{S}_k | T_k, \omega_k, \Theta) \\ &= \prod_{i=1}^{N_s(k)} \frac{1}{N_m} \sum_{j=1}^{N_m} \mathbb{E}(\mathbf{H}_{ij})_k p(\mathbf{s}_{k,i} | \mathbf{m}_{k,j}, T_k). \end{aligned} \quad (7)$$

Prior probability of observation parameters $p(Q_k | \Theta)$: The probability of the transformation T_k is independent of the model parameters Θ and the current weights ω_k , and $p(T_k)$ is assumed to be constant. The weights ω_k are modeled as Gaussian distributed with standard deviation λ_p :

$$\begin{aligned} p(Q_k | \Theta) &= p(\omega_k | \Theta) \\ &= \prod_{p=1}^n ((2\pi)^{n/2} \lambda_p)^{-1} \exp(-\omega_{k,p}^2 / 2\lambda_p^2). \end{aligned} \quad (8)$$

Prior probability of the model $p(\Theta)$: In Hufnagel *et al.* [12], $p(\Theta)$ is assumed to be constant, meaning implausible topological changes are not penalized. By introducing a higher number of degrees of freedom (position plus feature information) to our model a regularization becomes advantageous. Therefore, we define $p(\Theta)$ with the help of a Markov random field, where the position values for one sampling point depend on its local neighbors. The idea is to favor variation modes that keep the distance between the initial sampling point positions of the initial mean model and the adapted model instances small. Then, analog to $p(\mathbf{s}_{k,i} | \mathbf{m}_{k,j}, T_k)$, we model the distances between the initial mean model and the adapted model instances as a Gaussian distribution. It should be noted that $p(\Theta)$ does not penalize absolute point position distances but only relative distances between neighbors to avoid extreme topology changes.

2.2.2 Model Generation

Given our optimization criterion, $C(Q, \Theta)$ is minimized to determine the optimal model parameters Θ and observation parameters Q for the training data set. Another advantage of employing probabilistic correspondences is that the images in the training data set does not have to be complete in the sense that all possible structures have to be included in each image. Therefore, we are able to build an entire model of partial training images, where the missing parts are simply defined by zero-value correspondences and therefore are ignored during the model generation (see. Sec. 3.2 and Fig. 7 for an example).

For the optimization of $C(Q, \Theta)$ we use an iterative EM-strategy, which alternates between optimization of Θ

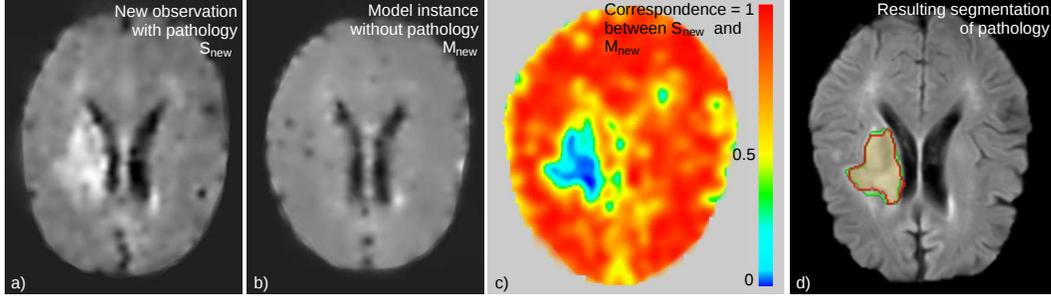


Figure 4. Classification of pathological structures not included in the model: if the mode is built employing only healthy images (e.g. MRI brain slices), the new input image (a) including a hyperintense pathological structure is represented by the model without the pathology (b). Low correspondence values between input \mathbf{S}_{new} and adapted \mathbf{M}_k (c) then indicate the location of the pathological structure. (d) shows the original input image with an expert segmentation (green) and the correspondence based segmentation of the pathology.

and updating the model-to-observation correspondences by computing T_k, ω_k for each observation (see Fig. 2). We can derive closed form solutions for all parameters except \mathbf{v}_p , which is computed with a Lagrange method subject to the constraints $\mathbf{v}_p^T \mathbf{v}_p = 1$ and $\mathbf{v}_p^T \mathbf{v}_q = 0$ for $q \neq p$ to ensure orthonormal variation modes. The most influential parameter of the generated model is the covariances matrix Σ in Eq. 1, determining which points are corresponding depending on their distances. We estimate Σ in each iteration based on the current distances between \mathbf{S}_k and \mathbf{M}_k .

2.3. Model Adaption

For the model adaption the model parameters Θ are fixed and $C(Q, \Theta)$ is minimized with respect to Q for a new unseen test image (see Fig. 2 step 4.-5.). Due to the vectorial representation of the image data our PAM is very flexible in terms of choice and number of appearance features with can describe the image information as well as any additional information. Furthermore, the use of probabilistic correspondences provide additional information about the adapted model instance. Therefore, we can apply our model to segmentation as well as classification.

2.3.1 Segmentation

To apply the model for segmentation purposes, additional information about given segmentations or landmarks of the training set is needed, which can be included into the model by adding further features describing a set of D_c contours/landmarks (C/L). We use distance maps for each of the given structures to ensure that all the sampling points contain information about their positions. Since there is no limit to the number of features, multi-object segmentation is included, where an extra distance feature is added for each object. Now a sampling point consists of position, features and segmentation information: $\mathbf{s}_i = (\mathbf{x}_i, \mathbf{f}_i, \mathbf{c}_i) \in \mathbb{R}^{(D_x + D_f + D_c)}$ with $\mathbf{c}_i = (c_{i,1}, \dots, c_{i,D_c})$. During the model generation process the segmentation information is

treated the same way as the appearance features. To adapt the model to a new and unseen image \mathbf{S}_{new} without the segmentation information, the computation of ω_{new} and T_{new} is done with a sub-vector of $\bar{\mathbf{M}}$ without contour information \mathbf{c}_i . The determined weights ω_{new} can then be used to reconstruct a model instance \mathbf{M}_{new} with contour information that is interpreted as the segmentation result for \mathbf{S}_{new} . Fig. 3 shows the segmentation process for a model with contour as well as label information (see Sec. 3.1), and furthermore, the model is built with additional points lying on the object's contour (green dots). Therefore, these points are in the model (model instances) but not in the new unseen images (without known contour). Due to the use of probabilistic correspondences, this does not lead to problems, but can be used to increase the accuracy of the model.

2.3.2 Classification

A pixelwise classification of medical image objects can also be considered as a multi-object segmentation employing label information for each class. But classifying pathologies, the considered structures can be highly heterogeneous which leads to problems using shape or appearance models to model these pathologies. Because models combining position and appearance information would need a training data set including all possible manifestations and locations of the considered structures, which is not practical when it comes to highly heterogeneous pathologies like e.g. stroke or tumor tissue in the brain. Another possibility is to build a model of the healthy/normal structure. Due to our use of probabilistic correspondences a "healthy" model can be adapted to new images including pathologies or occlusions. Then, the resulting correspondence probabilities indicate the regions where the new image includes "unhealthy" structures that do not fully correspond to the healthy model. In contrast to the segmentation process, the wanted information is not modeled as additional features but is given by the resulting correspondence proba-

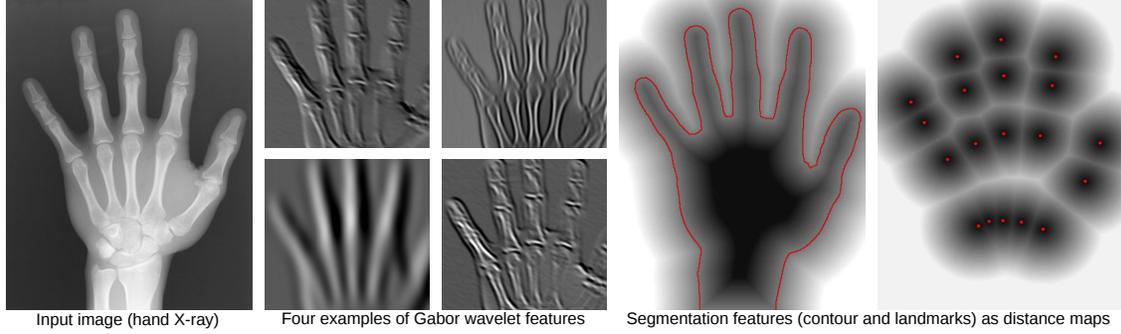


Figure 5. An example of the input data for a segmentation problem: 2D X-ray image of the hand with Gabor wavelet features as appearance features and hand contour and landmarks between finger joints as segmentation features described by distance maps.

bilities between the “unhealthy” test image and the adapted “healthy” model instance. We define for each observation point s_i a correspondence value by:

$$co_i = \sum_{j=1}^{N_m} p(s_i | \mathbf{m}_j, \theta). \quad (9)$$

Fig. 4 shows an example of a new observation with a pathology, its adapted model instance and the correspondence values between them.

3. Experiments

3.1. Segmentation

For the evaluation of the segmentation process the model is applied to X-ray 2D hand data. We use 20 image data sets from the “Digital Hand Atlas Database System”¹ as training data (10 × male, 10 × female, ASI, age 18). We evaluate the model adaption with 20 images (10 × male, 10 × female, ASI, age 17). For all 40 images I_k (with resolution of $0.5 \times 0.5 \text{ mm}^2$), we have the hand contour \mathcal{C}_k and an image with 19 landmarks of finger bone joints \mathcal{L}_k . As appearances we use 24 Gabor wavelets-filtered images (3 orientations at 4 scales each real/imaginary) and a sampling point spacing of 5 mm , leading to ≈ 1300 sampling points ($< 1\%$ of the input image pixel positions). In total, we have 28 dimensional vectors: 2 positions, 24 appearance and 2 segmentation features. We compute 10 variation modes.

To evaluate the modeled segmentation information we reconstruct the full contour/landmarks ($\mathcal{C}_k^{rec}, \mathcal{L}_k^{rec}$) from the observation \mathbf{S}_k . For \mathcal{C}_k^{rec} a scattered data B-spline interpolation is used as proposed by Lee *et al.* [14], where the sparse contour information (as distance map values) are the scattered data values. And for \mathcal{L}_k^{rec} a Hough circle transformation is employed, where the sampling position is the circle center and the distance map value of the associated

¹<http://www.ipilab.org/BAAWeb>

	MSCD (<i>mm</i>)	MPD (<i>mm</i>)
Sparse Repr.		
<i>input</i> vs. \mathbf{S}_k	0.06 (± 0.01)	0.23 (± 0.05)
Model Gene.		
\mathbf{S}_k vs. $\bar{\mathbf{M}}$	5.86 (± 1.02)	10.05 (± 2.71)
\mathbf{S}_k vs. \mathbf{M}_k	0.91 (± 0.18)	1.46 (± 0.37)
Segmentation		
<i>GT</i> vs. \mathbf{M}_{new}	1.64 (± 0.39)	2.42 (± 0.65)

Table 1. The table shows the results for the segmentation experiment 1-3 with mean symmetric contour distance (MSCD) for hand contours and the mean point distance (MPD) for joint landmarks.

joint feature defines the radius of the circle. The joint landmarks are then set on the locations of the local maxima. As evaluation measures, we compute the mean symmetric contour distance (MSCD) for hand contours and the mean point distance (MPD) for landmarks.

Experiment 1: Sparse Representation Accuracy: We evaluate the information loss from the input information (*input* with $\mathcal{C}_k, \mathcal{L}_k$) to the reconstructed point-based sparse representation (\mathbf{S}_k with $\mathcal{C}_k^{rec}, \mathcal{L}_k^{rec}$).

Experiment 2: Model Generation Accuracy: We compare reconstructed \mathcal{C}_k^{rec} and \mathcal{L}_k^{rec} of the 20 training observations \mathbf{S}_k with their corresponding model instances \mathbf{M}_k , to evaluate the accuracy of the model. For comparison, the mean symmetric contour distance and the mean point distance between \mathbf{S}_k and the unadapted model $\bar{\mathbf{M}}$ are computed.

Experiment 3: Model Adaption/Segmentation Accuracy: We adapt the model to 20 unseen test images without segmentation information and compare the resulting reconstructed segmentations of the model instances ($\mathbf{C}_{new}^{rec}, \mathbf{J}_{new}^{rec}$) with the given ground-truth segmentations *GT*.

See Tab. 1 for the results of the experiments. The accuracy of the sparse representation shows that a reconstruction leads to an error less than the used image spacing of 0.5 mm^2 for a sampling point spacing of 5 mm . By decreasing

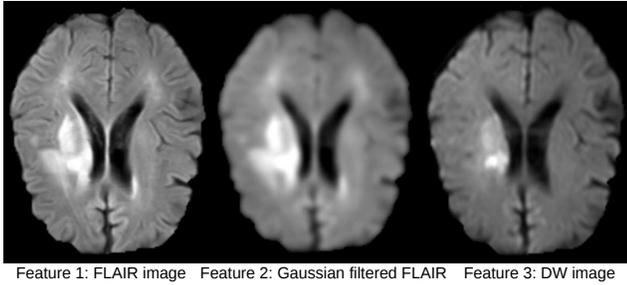


Figure 6. An example of the input data for a classification problem: a 2D-MRI slice from the middle of the brain of stroke patients with 3 features (FLAIR, smoothed FLAIR, DWI). In contrast to the segmentation problem of the hand, we do not have additional label features modeling the location of the stroke tissue.

the sampling point spacing further, the error becomes negligible. The evaluated model built with 20 training images and 10 variation modes is able to fit the 20 input data with an accuracy of 0.91 mm showing the ability of the model to represent the information in the training images. The segmentation results for 20 unseen images (1.64 mm for the contour and 2.42 mm for landmarks) reveal the capability of our approach to adapt and reconstruct contour/landmark information for unsegmented images.

3.2. Classification

For a proof of concept evaluation of the classification variant we use 2D slices of multi-spectral Magnetic Resonance Imaging (MRI) brain data of 44 patients with strokes (see Fig. 6) provided by the ISLES Challenge². The model is built with images from 22 patients and is adapted to 22 new unseen patient data. As appearance features we use the signal values of the “fluid-attenuated inversion recovery” (FLAIR) MRI sequence, the Gaussian-filtered FLAIR sequence and the diffusion weighted image (DWI), because these MRI-sequences show pathological brain regions. In total, we have 5 dimensional vectors: 2 positions, 3 appearance features and we compute 8 variation modes with a sampling point spacing of 5 mm .

We built the model using only the healthy regions of the 22 patient data. Employing correspondence probabilities a model of the entire structure can be built with partial image data without any further preprocessing steps. Fig. 7 shows examples of incomplete training data and the resulting model of the entire brain slice without stroke tissue.

To evaluate the classification between healthy and pathological tissue we reconstruct a full correspondence map from the scattered co_i values for each observation point s_i . Again a scattered data B-spline interpolation is employed and Fig. 4 shows an example of a reconstructed map, where highly corresponding regions are marked yellow-red

²<http://www.isles-challenge.org/>

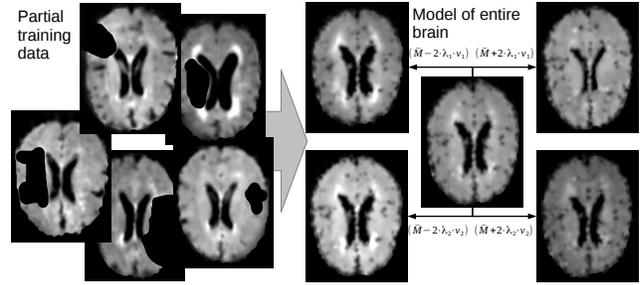


Figure 7. Model built with partial input training data with stroke tissue cut out: Only the first FLAIR-feature is visualized, reconstructed from the sparse image representations using a scattered data B-spline interpolation [14]. On the right the healthy model is visualized by the mean model and the first two variation modes.

and less corresponding regions are blue. For the evaluation we compared the thresholded reconstructed correspondence map with a given expert segmentation of the stroke affected areas. Tab. 2 shows the Dice’s coefficient, specificity and sensitivity for 3 different thresholds (50%, 37.5% and 25% correspondence). The correspondence maps indicate not only pathological regions but all variations not covered by the model. To identify connected regions with the low correspondences, we used a region growing algorithms with very low corresponding regions ($< 10\%$; dark blue in Fig. 4) as automatic seed points. Tab. 2 row 3 to 6 shows the results for 3 different thresholds for the region growing segmentation (50%, 37.5% and 25% correspondence). Fig. 8 shows three examples of correspondence based classifications.

The results (Tab. 2) show that the correspondence values are able to indicate the location of a pathology in unseen images. Due to the rather large sampling points spacing of the model (5 mm) the classification of small lesions is problematic. Therefore, the values enclosed in brackets show the results with excluded small lesions (diameter less than 5 mm) for comparison. The best Dice’s coefficients (0.6 for all and 0.68 for stroke regions over 5 mm) are achieved with a threshold of 37.5% correspondence with the region growing approach. The sensitivity obviously increases with a higher threshold leading to more tissue being classified as pathological and the specificity shows the best results for small threshold values. For comparison, other methods for stroke segmentation show Dice’s coefficients of ≈ 0.50 to 0.65 for 3D data [16].

4. Discussion

In this work we described a probabilistic approach for statistical appearance models in a MAP framework, where the model generation and the model fitting can be expressed as a single optimization criterion. A point-based image representation through a set of sampling points combining po-

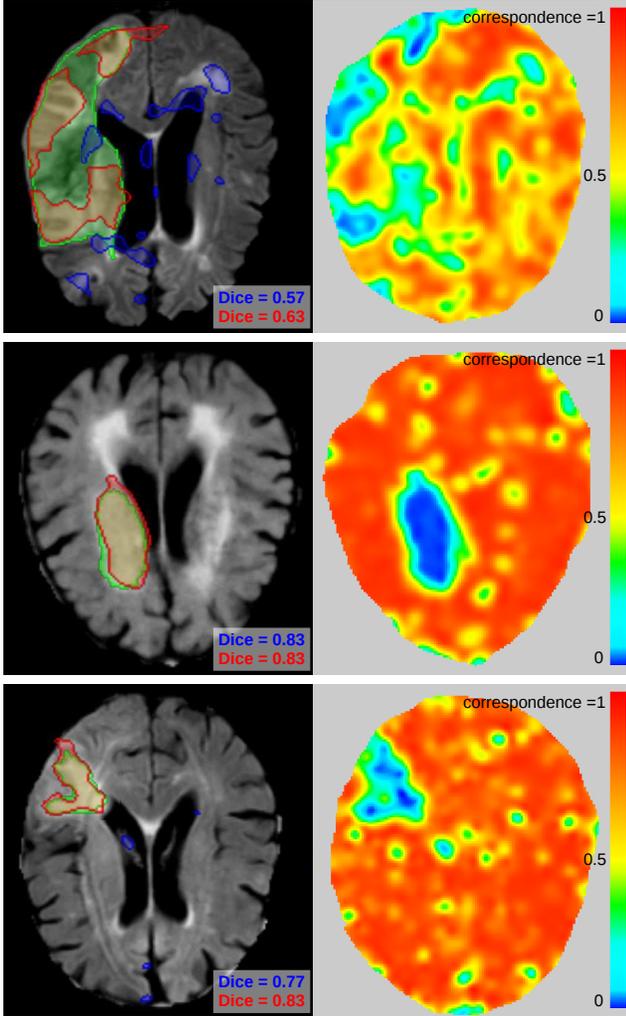


Figure 8. Three examples of classification results: of the left the input brain slide is shown with the expert segmentation (green) and the correspondence based segmentation (region growing: red; threshold red + blue). On the right the correspondence map between input and adapted model instance is visualized.

sition and appearance is used, modeling the shape information implicitly. In contrast to classical SSMs and AAMs, the use of probabilistic correspondences eliminates the costly predefinition of one-to-one correspondences and reduces the dependence of the generated model on the landmark positions. Furthermore, a model regularization is naturally included in the framework by using non-constant prior probabilities of the model.

We applied the model to model-based image segmentation, with multi-object segmentation included in the framework. The vectorial representation of the image data is very flexible in terms of choice of appearance features and additional information, which are not limited to contour (label) information but can contain any information that can be rep-

Corr.	Dice's Coeff.	Sensitivity	Specificity
0.25	0.50 (0.62)	0.57 (0.69)	0.51 (0.63)
0.375	0.47 (0.57)	0.71 (0.82)	0.40 (0.49)
0.5	0.39 (0.47)	0.80 (0.90)	0.29 (0.35)
0.25 (RG)	0.58 (0.71)	0.56 (0.69)	0.66 (0.81)
0.375 (RG)	0.60 (0.68)	0.71 (0.81)	0.57 (0.63)
0.5 (RG)	0.55 (0.61)	0.78 (0.88)	0.47 (0.51)

Table 2. Classification results: Dice's coefficient, sensitivity and specificity between expert segmentation of 22 unseen images and the correspondence based classification between healthy and pathological tissue. We compared three different thresholds for the correspondence maps with and without region growing (RG). The values enclosed in brackets show the results with excluded small lesions (diameter less than 5 mm) for comparison.

resented as a feature value. Therefore, the features can be selected application-specific to generate the most suitable data representation for the given task. The number of sampling points is also flexible and can be determined for each image individually.

In a proof-of-concept evaluation of the segmentation, we have shown the feasibility of the proposed approach and evaluated the model generation and model-based segmentation using 2D hand X-rays. The segmentation experiments reveal the capability of our approach to adapt and reconstruct contour/landmark information for unsegmented images. To improve the accuracy of the joint detection, the training data size have to be increased and more joint specific features can be included into the feature vector. Employing correspondences probabilities instead of one-to-one correspondences not only eliminates the need of expensive preprocessing but also provides additional information about the local matching properties between unseen test image and adapted model instance. We use these information for classification purposes between healthy and pathological brain tissue in MRI slices. First results show the ability of the model to distinguish between healthy tissue (included in the model) and pathological tissue (unknown by the model). The problems detecting smaller lesions can simply be solved by decreasing the sampling point spacing in further evaluations. Furthermore, we show that using correspondences probabilities a model of an entire structure can be built using partial training images.

The proposed approach describes the modeling process in a concise and flexible mathematical framework. This opens up new directions for statistical appearance models and lays the foundation for further developments.

Acknowledgment: This work is supported by the German Research Foundation (DFG: HA 2355/7-2).

References

- [1] A. Andreopoulos and J. K. Tsotsos. Efficient and generalizable statistical models of shape and appearance for analysis of cardiac MRI. *Medical Image Analysis*, 12(3):335 – 357, 2008. [1](#)
- [2] H. Chui and A. Rangarajan. A new point matching algorithm for non-rigid registration. *Computer Vision and Image Understanding*, 89(23):114 – 141, 2003. [2](#)
- [3] T. Cootes, C. Taylor, D. Cooper, and J. Graham. Active shape models—their training and application. *Computer Vision and Image Understanding*, 61(1):38 – 59, 1995. [1](#)
- [4] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 23(6):681–685, June 2001. [2](#)
- [5] D. Cristinacce and T. Cootes. Feature detection and tracking with constrained local models. pages 929–938, 2006. [2](#)
- [6] R. Donner, G. Langs, and H. Bischof. Sparse mrf appearance models for fast anatomical structure localisation. In *In British Machine Vision Conference*, 2007. [2](#)
- [7] J. Ehrhardt, J. Krüger, and H. Handels. Statistical shape and appearance models without one-to-one correspondences. In S. Ourselin and M. Styner, editors, *SPIE Medical Imaging 2014, Image Processing*, volume 9034, page 90340U, San Diego, USA, 2014. [2](#)
- [8] S. Gold, A. Rangarajan, C. ping Lu, and E. Mjolsness. New algorithms for 2D and 3D point matching: Pose estimation and correspondence. *Pattern Recognition*, 31:957–964, 1997. [2](#)
- [9] S. Granger and X. Pennec. Multi-scale EM-ICP: A fast and robust approach for surface registration. In *Computer Vision ECCV 2002*, volume 2353 of *Lecture Notes in Computer Science*, pages 418–432. Springer, 2002. [2](#), [3](#)
- [10] T. Heimann and H.-P. Meinzer. Statistical shape models for 3D medical image segmentation: A review. *Medical Image Analysis*, 13(4):543 – 563, 2009. [1](#)
- [11] H. Hufnagel, X. Pennec, J. Ehrhardt, N. Ayache, and H. Handels. Generation of a statistical shape model with probabilistic point correspondences and the expectation maximization-iterative closest point algorithm. *International Journal of Computer Assisted Radiology and Surgery*, 2:265 – 273, 3/2008 2008. [2](#)
- [12] H. Hufnagel, X. Pennec, J. Ehrhardt, H. Handels, and N. Ayache. Shape analysis using a point-based statistical shape model build on correspondence probabilities. In *10th International Conference on Medical Image Computing and Computer-Assisted Intervention, MICCAI*, volume Part I, pages 959 – 967. Springer, 2007. [2](#), [4](#)
- [13] B. Jian and B. C. Vemuri. Robust point set registration using gaussian mixture models. *IEEE Trans. Pattern Anal. Mach. Intell.*, 33(8):1633–1645, Aug. 2011. [2](#)
- [14] S. Lee, G. Wolberg, and S. Y. Shin. Scattered data interpolation with multilevel B-splines. *IEEE Trans on Visualization and Computer Graphics*, 3:228–244, 1997. [6](#), [7](#)
- [15] T. S. Lee. Image representation using 2D gabor wavelets. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 18:959–971, 1996. [2](#)
- [16] O. Maier, M. Wilms, J. von der Gablentz, U. M. Krmer, T. F. Mnte, and H. Handels. Extra tree forests for sub-acute ischemic stroke lesion segmentation in MR sequences. *Journal of Neuroscience Methods*, 240(0):89 – 100, 2015. [7](#)
- [17] S. Mitchell, J. Bosch, B. P. F. Lelieveldt, R. van der Geest, J. Reiber, and M. Sonka. 3-D active appearance models: segmentation of cardiac MR and ultrasound images. *Medical Imaging, IEEE Transactions on*, 21(9):1167–1178, 2002. [1](#)
- [18] A. Myronenko and X. Song. Point set registration: Coherent point drift. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32(12):2262–2275, Dec 2010. [2](#)
- [19] H. Xu, B. Thirion, and S. Allassonnire. Probabilistic atlas and geometric variability estimation to drive tissue segmentation. *Statistics in Medicine*, 33(20):3576–3599, 2014. [2](#)
- [20] P. Zhang and T. F. Cootes. Automatic construction of parts+geometry models for initializing groupwise registration. *IEEE Trans. Med. Imaging*, 31(2):341–358, 2012. [1](#)