Accurate and Robust Registration of Nonrigid Surface using Hierarchical Statistical Shape Model

Hidekata Hontani Yuto Tsunekawa Yoshihide Sawada
Nagoya Institute of Technology
Aichi, JAPAN
hontani@nitech.ac.com

Abstract

In this paper, we propose a new non-rigid robust registration method that registers a point distribution model (PDM) of a surface to given 3D images. The contributions of the paper are (1) a new hierarchical statistical shape model (SSM) of the surface that has better generalization ability is introduced, (2) the registration algorithm of the hierarchical SSM that can estimate the marginal posterior distribution of the surface location is proposed, and (3) the registration performance is improved by (3-1) robustly registering each local shape of the surface with the sparsity regularization and by (3-2) referring to the appearance between the neighboring model points in the likelihood computation. The SSM of a liver was constructed from a set of clinical CT images, and the performance of the proposed method was evaluated. Experimental results demonstrated that the proposed method outperformed some existing methods that use non-hierarchical SSMs.

1. Introduction

Statistical shape models (SSMs) of surfaces have proved to be an important basis for 3D image segmentation[8]. Given an SSM of a target surface, one can segment the target regions in given 3D images by registering the SSM to them. This paper assumes that 3D medical images are input and that the objective of the registration is to segment anatomical structures in the images. Medical image segmentation is not an easy task because, in general, there exist many nuisances[12], which impede identifying the boundary of the target organ: Many false boundaries are often detected from the organs other than the target one, and some parts of the true boundary can be missed because of lesions or of contacts with the neighboring organs. SSMs and their robust registration techniques are both required for accurately determining the boundaries of target organs against those nuisances.

In this paper, a point distribution model (PDM) is employed: A set of N points represents the surface. Let the 3D coordinates of the i-th point be denoted by $\boldsymbol{x}_i$ $(i = 1, 2, \cdots, N)$. Then, the surface is represented by a $3N$-vector, $\boldsymbol{X} = [\boldsymbol{x}_1^T, \boldsymbol{x}_2^T, \cdots, \boldsymbol{x}_N^T]^T$. Let a set of the training surfaces, of which locations and sizes are normalized in advance, be denoted by S^l $(l = 1, 2, \cdots, M)$, and a set of N corresponding points generated on S^l be denoted by $\boldsymbol{x}_i^l$ $(i = 1, 2, \cdots, N,\ l = 1, 2, \cdots, M)$. Let a $3N$-vector, $\boldsymbol{X}^l = [(\boldsymbol{x}_1^l)^T, (\boldsymbol{x}_2^l)^T, \cdots, (\boldsymbol{x}_N^l)^T]^T$, denote the distribution of the points on S^l. The SSM for a PDM is constructed from this training set, $\mathcal{X} = \{\boldsymbol{X}^l | l = 1, 2, \cdots, M\}$.

1.1. Generalization Ability of SSM

One of the problems in constructing SSMs for anatomical structures is that it is not easy to collect enough number of training data. It should be reminded that each training surface is obtained by labeling the boundary of the target organ in a 3D medical image. Manual operation by an expert is required for this labeling. Davies et al. introduced the following criteria for comparing SSMs[4]: (1)Generalization ability, (2) Specificity, and (3) Compactness. The generalization ability measures the ability to describe any instance of the target shapes – not just those seen in the training set. The specificity, on the other hand, measures the restriction of the representation: The SSM should be able to describe only valid instances of the target shapes. These criteria are employed for the comparison of SSMs. When the number of the training data is small, the generalization ability is more seriously required for the SSMs.

This paper concerns itself with linear SSMs of surfaces. An active shape model (ASM)[2] is one of the most important linear SSMs used for nonrigid registration. The original ASM represents the shape variety of a target surface with a subspace, which is spanned by a set of principal eigenvectors of the covariance matrix, $\Sigma_{\mathcal{X}}$, of $\{\boldsymbol{X}^l | l = 1, 2, \cdots, M\}$. The set of eigenvectors is obtained by applying PCA to $\Sigma_{\mathcal{X}}$, and those with smaller eigenvalues are

1063-6919/13 $26.00 © 2013 IEEE
DOI 10.1109/CVPR.2013.383

excluded from the bases of the subspace. This exclusion improves the specificity and the compactness of the ASM but makes the generalization ability worse. When the number of the training data is small, this drawback of the lack of the generalization ability is serious, hence many methods have been proposed for improving the generalization ability[3][11][14].

One of the approaches for the improvement is to make the ASM hierarchical[3][14]. Davatzikos et al. described the goal of the hierarchical scheme as utilizing marginal distribution that can be estimated accurately[3]. As mentioned above, the SSM of PDM represents the surface with $3N$ variables. Among these variables, the hierarchical scheme firstly examines smaller sets of highly correlated variables, and then, the joint distribution of a large number of variables is considered after smoothing, which makes the distribution more compact[3]. The proposed method follows this scheme. Two strategies are proposed in [3] for the scheme. One is to partition the surface into several segments, and the other is to apply wavelet transformation to $\boldsymbol{X}$. In the former strategy, one PCA-based SSM is constructed for representing the shape of each segment, and one more PCA-based SSM is constructed for representing the locations of all of the segments. In this paper, we employ this partitioning approach for making the SSM hierarchical, because this approach is appropriate for introducing a robust registration technique described below.

Another approach for improving the generalization ability is to make the structure of the corresponding graphical model sparse[11]. A linear SSM can be represented by a Gaussian graphical model, and the marginal posterior probability distribution of $\boldsymbol{X}$ can be inferred on the graphical model by means of, for example, a belief propagation or of MCMC[10]. In the graph, each node represents $\boldsymbol{x}_i$, and two nodes are linked by an edge if the two corresponding points, $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ are conditionally dependent. It should be reminded that the structure of the Gaussian graphical model is determined by the precision matrix, which is the inverse of the covariance: Two points, $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ are conditionally dependent if and only if the (i, j) component of the precision matrix is not zero. Let the empirical covariance matrix obtained from the training data be denoted by $\Sigma_{\mathcal{X}}$, and let the corresponding precision matrix be denoted by $\Lambda_{\mathcal{X}} = \Sigma_{\mathcal{X}}^{-}$. In many cases, all components of $\Lambda_{\mathcal{X}}$ are non-zero. As pointed out in [6], an empirical covariance matrix, $\Sigma_{\mathcal{X}}$, can over-fit to the training data, when the number of the data is small. This is also one of the main reasons why the generalization ability of ASM is poor. Friedman et al. proposed a graphical lasso, which is a method for estimating a precision matrix from the empirical covariance matrix using a sparsity regularization[6]. Using the graphical lasso, one can obtain a sparse Gaussian graphical model from the training set. It has been reported that, one can improve the registration accuracy by estimating the structure of the Gaussian graphical model of the SSM by means of the graphical lasso[11]. We represent the global SSM using a graphical model of which structure is determined by means of the graphical lasso, in the proposed method.

1.2. Robust Registration

Robust registration methods are needed for accurately registering SSMs to given medical images because of the nuisances mentioned above. The nuisances are the outliers of the statistical models and should be excluded for the accurate registration. Many robust registration methods have been proposed[7][9][10][15]. Among them, some methods use a sparsity regularization techniques for the robust nonrigid registration[10][15]. Let a target point in a given image corresponding to $\boldsymbol{x}_i$ be denoted by $\boldsymbol{y}_i$, and let $\boldsymbol{Y} = [\boldsymbol{y}_1^T, \boldsymbol{y}_2^T, \cdots, \boldsymbol{y}_N^T]^T$. Assume that $\boldsymbol{X}$ is represented by a function of a set of parameters, $\boldsymbol{\theta}$: $\boldsymbol{X} = \boldsymbol{X}(\boldsymbol{\theta})$. Then, the minus log likelihood functions in many non-robust non-rigid registration methods can be simply represented as follows: $E'(\boldsymbol{\theta}) = \|\boldsymbol{Y} - \boldsymbol{X}(\boldsymbol{\theta})\|^2$. The robust registration methods based on the sparsity regularization introduce a $3N$-vector, $\boldsymbol{E} = [\boldsymbol{e}_1^T, \boldsymbol{e}_2^T, \cdots, \boldsymbol{e}_N^T]^T$, where the three-vector, $\boldsymbol{e}_i$, becomes non-zero if $\boldsymbol{y}_i$ is an outlier of the SSM[10][15]. Those methods minimize the following cost function: $E(\boldsymbol{\theta}, \boldsymbol{E}) = \|\boldsymbol{Y} - \boldsymbol{X}(\boldsymbol{\theta}) - \boldsymbol{E}\|^2 + \rho|\boldsymbol{E}|_{L1}$ with respect to $\boldsymbol{\theta}$ and $\boldsymbol{E}$. If $\boldsymbol{E}$ is not penalized by the L1-norm term, you can always make the value of $\|\boldsymbol{Y} - \boldsymbol{X}(\boldsymbol{\theta}) - \boldsymbol{E}\|^2$ zero by setting $\boldsymbol{E} = \boldsymbol{Y} - \boldsymbol{X}(\boldsymbol{\theta})$. Because $\boldsymbol{E}$ is penalized by its L1-norm, the resultant vector, $\boldsymbol{E}$, often becomes sparse and $\boldsymbol{e}_i$ become non-zero only when $|\boldsymbol{y}_i - \boldsymbol{x}_i|$ are large. Let Ω denote the outlier regions in which $\boldsymbol{y}_i$ are the outliers. Then, $\|\boldsymbol{Y} - \boldsymbol{X}(\boldsymbol{\theta}) - \boldsymbol{E}\|^2 = \sum_{\boldsymbol{y}_i \notin \Omega} \|\boldsymbol{y}_i - \boldsymbol{x}_i\|^2$, because $\boldsymbol{e}_i = \boldsymbol{y}_i - \boldsymbol{x}_i$ in Ω at the stationary point of E. In other words, minimizing $E(\boldsymbol{\theta}, \boldsymbol{E})$, one can detect outliers and can exclude them to estimate the parameter, $\boldsymbol{\theta}$, accurately.

It should be noted that outliers can be appropriately detected only when the model has enough specificity: Otherwise, model cannot distinguish between the outliers and inliers. We employ PCA-based SSM for representing the shape of each segment because such the model has high ability for the specificity, as mentioned above. Representing the local shape of each segment by a PCA-based SSM, we introduce the sparsity regularization for registering each segment robustly.

1.3. Likelihood of Point Location

A set of candidate points of the boundary of the target organ should be firstly detected. Then, each target point, $\boldsymbol{y}_i$, is located at some of those candidate points. The candidate points are detected based on the likelihood of $\boldsymbol{x}_i$: The likelihood function, $L(\boldsymbol{x}_i) = p(I|\boldsymbol{x}_i)$, is designed so that the likelihood becomes high on the true boundary to be

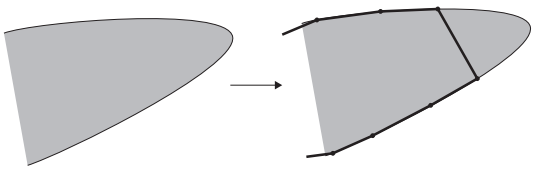

Figure 1. An example of registration error that exists even after all model points are located on the true boundary

extracted from the image. For example, $L(\boldsymbol{x}_i) = \delta(\Delta I)$ for any i, if the candidate points are detected by applying a zero-crossing detector to the Laplacian of the image. In most of all methods of PDM registration, the likelihood of $\boldsymbol{x}_i$ is determined based only on the local appearance, I_i, around $\boldsymbol{x}_i$: $L(\boldsymbol{x}_i) = p(I_i|\boldsymbol{x}_i)$. This often causes large registration errors. Firstly, the target points can be located on false boundaries in a given image, because one cannot distinguish between the true boundary and the false ones based only on the local appearances. Secondly, the continuity of the boundary in a given image is not evaluated, and hence there can exist large residuals between the registered surface and the true boundary in the image even when all points of the model are located on the true boundary, as shown in Fig.1. In order to reduce these residuals, you need to determine the locations of the model points based not only on the local appearance around each point but also on the appearance *between* neighboring points.

In this paper, we determine the likelihood based on the appearance around a line segment, s_{ij}, that connects two neighboring points, $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$. As the result, the likelihood is represented by a two-variable function, $L(\boldsymbol{x}_i, \boldsymbol{x}_j) = p(I_i|\boldsymbol{x}_i)p(I_j|\boldsymbol{x}_j)p(I_{ij}|\boldsymbol{x}_i, \boldsymbol{x}_j)$, where I_{ij} is the local appearance around s_{ij}. The arguments of the likelihood are the pairwise variables, and the computation of the likelihood can be incorporated into the registration algorithms by representing the SSM using a pairwise graphical model[5].

1.4. Contributions

The contributions presented in this paper are the followings. (1) A new hierarchical SSM that has better generalization ability is introduced, (2) the registration algorithm of the hierarchical SSM that can estimate the marginal posterior distribution is proposed, and (3) the registration performance is improved by robustly registering each segment with the sparsity regularization and by referring to the appearance between the neighboring points in the likelihood computation.

2. Method

In this section, the proposed hierarchical SSM is firstly described, and then the registration algorithm is explained.

2.1. Construction of Hierarchical SSM

In the proposed method, a target surface is partitioned into K segments, a linear SSM is constructed for each of the segments, and the relationship between the shapes of the segments is represented by a Gaussian graphical model. The models are constructed from a set of training surfaces. Let a set of training 3D images be denoted by $\{I^l|l = 1, 2, \cdots, M\}$, where M is the number of the images. Assume that the location, size and the shape of the bodies in the images are normalized based on a set of landmarks around the target organ. Let the surface of a target organ manually labeled in I^l be denoted by S^l. A set of N corresponding points is generated on S^l. Let the i-th corresponding point on S^l be denoted by $\boldsymbol{x}_i^l$, and let $\bar{\boldsymbol{x}}_i$ denote the average of $\{\boldsymbol{x}_i^l|l = 1, 2, \cdots, M\}$.

2.1.1 Statistical Shape Model for Each Segment

The surface is partitioned into K segments by applying, e.g., K-means method to $\{\bar{\boldsymbol{x}}_i|i = 1, 2, \cdots, N\}$. Let a set of the indexes of the points included in the k-th segment be denoted by $\mathcal{L}^k = \{i_1^k, i_2^k, \cdots, i_{n(k)}^k\}$, where $n(k)$ is the number of points included in the segment. Let $\mathcal{P}^k = \{\boldsymbol{x}_i^l|i \in \mathcal{L}^k, l = 1, 2, \cdots, M\}$.

We employ a probabilistic PCA (PPCA)[13] for the model construction, because PPCA is derived from a probabilistic model explicitly including a model for observation, and because the number of free parameters to be estimated in PPCA is smaller than that in PCA. The former characteristic is needed for representing the statistical relationship between the model points and the shape parameters, and the latter one is a strength when the number of training samples is small. Let $\boldsymbol{X}^k = [\boldsymbol{x}_{i_1^k}^T, \boldsymbol{x}_{i_2^k}^T, \cdots, \boldsymbol{x}_{i_{n(k)}^k}^T]^T$. The PPCA assumes the following linear model:

$$\boldsymbol{X}^k = W^k\boldsymbol{\theta}^k + \bar{\boldsymbol{X}}^k + \boldsymbol{\epsilon}, \tag{1}$$

where W^k is a $3n(k) \times q^k$ matrix ($3n(k) > q^k$), $\bar{\boldsymbol{X}}^k$ is the mean vector, and $\boldsymbol{\epsilon}$ denotes an isotropic zero-mean Gaussian noise:

$$\boldsymbol{\epsilon} \sim \mathcal{N}(0, (\sigma^k)^2 I), \tag{2}$$

where I is an identity matrix and $(\sigma^k)^2$ is the variance. In PPCA, one estimates W^k, $\bar{\boldsymbol{X}}^k$, and $(\sigma^k)^2$ by using the EM algorithm. The conditional probability distribution of $\boldsymbol{\theta}^k$ is represented by a Gaussian as follows:

$$p(\boldsymbol{\theta}^k|\boldsymbol{X}^k) = \mathcal{N}((M^k)^{-1}(W^k)^T(\boldsymbol{X}^k - \bar{\boldsymbol{X}}^k), (\sigma^k)^2(M^k)^{-1}), \tag{3}$$

where $M^k = (W^k)^T(W^k) + (\sigma^k)^2 I$. Exactly speaking, in our settings, $\boldsymbol{\theta}^k$ represents not only the shape but also the location and the orientation of the k-th segment. $\boldsymbol{\theta}^k$ obeys a Gaussian:

$$p(\boldsymbol{\theta}^k) = \mathcal{N}(\boldsymbol{0}, I), \tag{4}$$

and $\boldsymbol{X}^k$ obeys a Gaussian as follows:

$$p(\boldsymbol{X}^k|\boldsymbol{\theta}^k) = \mathcal{N}(W^k\boldsymbol{\theta}^k + \bar{\boldsymbol{X}}^k, (\sigma^k)^2 I). \quad (5)$$

Given a data, $\boldsymbol{Y}^k$, one can obtain the maximum likelihood estimate of the shape parameter, $\boldsymbol{\theta}^k$, by minimizing

$$\Delta^k(\boldsymbol{\theta}^k) = \|\boldsymbol{Y}^k - W^k\boldsymbol{\theta}^k - \bar{\boldsymbol{X}}^k\|^2. \quad (6)$$

The dimension of $\boldsymbol{\theta}^k$ is q^k, and is much smaller than the number of parameters needed to describe the locations of all $n(k)$ points in the segment. Each component of $\boldsymbol{\theta}^k$ corresponds to some deformation mode represented by each column of W^k.

2.1.2 Statistical Model for All Segments

A statistical model that represents the relationship between the shapes of the segments is then constructed. The model is represented by a Gaussian graphical model of $\{\boldsymbol{\theta}^k|k = 1, 2, \cdots, K\}$, and its structure is estimated by using a graphical lasso[6]. Let $\boldsymbol{\Theta} = [\boldsymbol{\theta}^{1T}, \boldsymbol{\theta}^{2T}, \cdots, \boldsymbol{\theta}^{KT}]^T$ represents the shape parameters of all segments. Let $\boldsymbol{\theta}^{k,l}$ denote the ML estimate of the shape parameter obtained from $\boldsymbol{X}^{k,l} = [\boldsymbol{x}_{i_1^k}^{lT}, \boldsymbol{x}_{i_2^k}^{lT}, \cdots, \boldsymbol{x}_{i_{n(k)}^k}^{lT}]^T$, and let $\boldsymbol{\Theta}^l = [(\boldsymbol{\theta}^{1,l})^T, (\boldsymbol{\theta}^{2,l})^T, \cdots, (\boldsymbol{\theta}^{K,l})^T]^T$, which is used as the training sample obtained from S^l. Let an empirical covariance matrix of $\{\boldsymbol{\Theta}^l|l = 1, 2, \cdots, M\}$, be denoted by Σ_Θ. The number of the training samples, M, can be smaller than the dimension of $\boldsymbol{\Theta}$, though the dimension of $\boldsymbol{\Theta}$ is much smaller than $3N$. We introduce the graphical lasso for handling the problem of the over-fitting of empirical covariance matrix.

In the graphical lasso, the precision matrix, Λ, is estimated as follows:

$$\hat{\Lambda} = \arg\min_{\Lambda}\{-\log\det\Lambda + \mathrm{tr}\,(\Sigma_\Theta)\Lambda) + \lambda\|\Lambda\|_{L1}\}, \quad (7)$$

where λ is a positive coefficient for the regularization. In the proposed model, $\boldsymbol{\Theta}$ is assumed to obey a Gaussian with its mean is zero and its precision matrix is $\hat{\Lambda}$ as follows:

$$p(\boldsymbol{\Theta}) = \mathcal{N}(\boldsymbol{0}, \hat{\Lambda}^-). \quad (8)$$

The mean of $\boldsymbol{\Theta}$ is zero because of (4). This Gaussian distribution is represented with a graphical model, of which edges correspond to the non-zero off diagonal components of $\hat{\Lambda}$. It should be noted that this graphical model represents the conditional dependencies between the deformation modes of different segments, and that the structure of the graph should be sparse, if the shape parameters of two distantly-located segments are conditionally independent. The SSM of the proposed method is now represented as follows:

$$p(\{\boldsymbol{X}^k\}, \{\boldsymbol{\theta}^k\}) = \prod_k p(\boldsymbol{X}^k|\boldsymbol{\theta}^k)p(\boldsymbol{\Theta}). \quad (9)$$

2.2. Likelihood of Target Point Location

For registering k-th segment, a target point is detected from the given image based on the likelihood, $L^k(\boldsymbol{X}^k) = L^k(\{\boldsymbol{x}_i|i \in \mathcal{L}^k\})$. It is defined as follows in the method:

$$L^k(\boldsymbol{X}^k) = \prod_{i\in\mathcal{L}^k} p(I_i|\boldsymbol{x}_i) \prod_{\substack{i,\,j\,\in\,\mathcal{L}^k,\\ (i,\,j)\,\in\,\mathcal{E}_\mathrm{D}}} p(I_{ij}|\boldsymbol{x}_i, \boldsymbol{x}_j), \quad (10)$$

where $\mathcal{E}_\mathrm{D}$ is a set of edges of a Delaunay triangulation of $\bar{\boldsymbol{x}}_i \in \mathcal{P}^k$ on the surface. Here, let I_i is a L^3-vector, which denotes the appearance in a $L \times L \times L$ cube of which center is at $\boldsymbol{x}_i$. Analogously, let I_i^l is a L^3-vector, which denotes the local appearance around $\boldsymbol{x}_i^l$ in I^l.

Assuming $p(I_i|\boldsymbol{x}_i)$ is a Gaussian, we represent it as follows:

$$p(I_i|\boldsymbol{x}_i) = \mathcal{N}(\bar{I}_i, \Sigma_{I_i}), \quad (11)$$

where the mean vector, $\bar{I}_i$, and the variance matrix, Σ_{I_i}, are both estimated based on the training set, $\{I_i^l|l = 1, 2, \cdots, M\}$. It should be reminded that the likelihood distribution is not necessarily a Gaussian, even though the conditional probability distribution is a Gaussian.

$p(I_{ij}|\boldsymbol{x}_i, \boldsymbol{x}_j)$ is introduced for evaluating the appearance between two model points. For efficient computation, we define it with the internally dividing points of s_{ij} as follows:

$$p(I_{ij}|\boldsymbol{x}_i, \boldsymbol{x}_j) = \prod_{t=1}^{T-1} p(I_{ij,t}|\boldsymbol{x}_{ij,t}). \quad (12)$$

Here, $T = \lfloor\|\boldsymbol{x}_i - \boldsymbol{x}_j\|/\Delta\rfloor$, where Δ is a constant determined in advance. The dividing points, $\boldsymbol{x}_{ij,t} = (t\boldsymbol{x}_i + (T-t)\boldsymbol{x})/T$ $(t = 1, 2, \cdots, T-1)$, are located at regular intervals, Δ, on the segment, s_{ij}. $I_{ij,t}$ is also a L^3-vector, which denotes the local appearance around $\boldsymbol{x}_{ij,t}$. $p(I_{ij,t}|\boldsymbol{x}_{ij,t})$ is assumed to obey a Gaussian, and its mean and covariance are estimated based on a set of the training data, $\{I_{ij,t}^l|l = 1, 2, \cdots, M\}$, which are the local appearance around $\boldsymbol{x}_{ij,t}^l$. It should be noted that incorporating the internally dividing points $\boldsymbol{x}_{ij,t}$, is different from just increasing the number of the model points, M. You cannot eliminate the residuals indicated in Fig.1 only by increasing M.

Incorporating the SSM in (9), we represent the simultaneous posterior distribution of the model points and the shape parameters as follows:

$$p(\{\boldsymbol{X}^k\}, \{\boldsymbol{\theta}^k\}|I) \propto \prod_k L(\boldsymbol{X}^k)p(\boldsymbol{X}^k|\boldsymbol{\theta}^k)p(\boldsymbol{\Theta}). \quad (13)$$

2.3. Hierarchical Registration Algorithm

In the proposed method, the marginal posterior distribution of the shape parameters, $p(\boldsymbol{\theta}^k|I)$, is estimated for each segment. Then, the marginal distribution of the model

points, $\boldsymbol{X}^k$, can be obtained by using (5). The proposed method estimates $p(\boldsymbol{\theta}^k|I)$ iteratively as follows:

1. Set $p^0(\boldsymbol{\theta}^k|I) = p(\boldsymbol{\theta}^k) = \mathcal{N}(0, I)$ as shown in (4).
2. For $\tau = 1, 2, \cdots, \mathcal{T}$
   (a) Estimate the temporal posterior distribution, $p_k^\tau(\boldsymbol{\theta}^k|I)$, for each segment by registering the local SSM of each segment to the given image. The robust registration technique with the sparsity regularization is used at this step.
   (b) Update the posterior distribution to obtain $p^\tau(\boldsymbol{\theta}^k|I)$ based on $p_k^\tau(\boldsymbol{\theta}^k|I)$ and on $p(\boldsymbol{\Theta})$ in (8).

The details of (a) and (b) are described in the followings.

2.3.1 Robust Registration of Local Segment Model

For estimating the temporal posterior distribution, $p_k^\tau(\boldsymbol{\theta}^k|I)$, for each segment, the target points, $\boldsymbol{Y}^k$, is firstly detected in the image, and then, the SSM is registered to the target points by estimating the posterior distribution. The target points are sampled from the given image based on the conditional probability distribution, $p(\boldsymbol{X}^k|I)$, which can be obtained by marginalizing $p(\boldsymbol{X}^k|I, \boldsymbol{\theta}^k)$ with respect to $\boldsymbol{\theta}^k$.

If an image and the shape parameter, $\boldsymbol{\theta}^k$, is given, the conditional distribution of the model points, $\boldsymbol{X}^k$ is given as follows:

$$p(\boldsymbol{X}^k|I, \boldsymbol{\theta}^k) \propto L^k(\boldsymbol{X}^k) p(\boldsymbol{X}^k|\boldsymbol{\theta}^k), \tag{14}$$

where $L^k(\cdot)$ is shown in (10) and $p(\boldsymbol{X}^k|\boldsymbol{\theta}^k)$ is a Gaussian as shown in (5). Assume that the posterior distribution of the shape parameter, $p^{\tau-1}(\boldsymbol{\theta}^k|I)$ is given. Then, the posterior distribution of the model points is obtained by marginalizing $p(\boldsymbol{X}^k|I, \boldsymbol{\theta}^k)$ in (14) with respect to $\boldsymbol{\theta}^k$ as follows:

$$p^{\tau-1}(\boldsymbol{X}^k|I) \propto L^k(\boldsymbol{X}^k) \int p(\boldsymbol{X}^k|\boldsymbol{\theta}^k) p^{\tau-1}(\boldsymbol{\theta}^k|I) d\boldsymbol{\theta}^k. \tag{15}$$

In the method, $p^{\tau-1}(\boldsymbol{\theta}^k|I)$ is represented non-parametrically with a set of samples, $\{\boldsymbol{\theta}_u^{k,\tau-1}|u = 1, 2, \cdots, U\}$, because its distribution is highly non-linear in many cases. Using this non-parametric representation, we approximate the integration in (15) as follows:

$$p^{\tau-1}(\boldsymbol{X}^k|I) \simeq \frac{1}{Z} L^k(\boldsymbol{X}^k) \sum_u p(\boldsymbol{X}^k|\boldsymbol{\theta}_u^{k,\tau-1}). \tag{16}$$

The method for making the non-parametric representation of $p^{\tau-1}(\boldsymbol{\theta}^k|I)$ will be described later.

Here, a set of target points, $\mathcal{Y}^{k,\tau-1} = \{\boldsymbol{Y}_u^{k,\tau-1}|u = 1, 2, \cdots, U\}$ is drawn from the distribution, $p^{\tau-1}(\boldsymbol{X}^k|I)$ in (16). It should be noted that $p(\boldsymbol{X}^k|I, \boldsymbol{\theta}^k)$ in (14) can be represented by a pairwise Markov random field, and that one can draw the samples from $p(\boldsymbol{X}^k|I, \boldsymbol{\theta}^k)$ by using some sampling technique, for example, a MCMC or a belief propagation. For each of the target points, $\boldsymbol{Y}_u^{k,\tau-1}$, the maximum likelihood estimate of the shape parameter is robustly computed based on the model shown in (5) with a sparsity regularization[10][15]. The cost function to be minimized is as follows:

$$E(\boldsymbol{\theta}^k, \boldsymbol{e}) = \|\boldsymbol{Y}_u^{k,\tau-1} - W^k\boldsymbol{\theta}^k - \bar{\boldsymbol{X}}^k - \boldsymbol{e}\|^2 + \rho\|\boldsymbol{e}\|_1, \tag{17}$$

where ρ is a positive coefficient for the regularization, and $\boldsymbol{e}$ denotes the residuals corresponding to the outliers. The cost function in (17) is convex with respect to $\boldsymbol{\theta}$ and $\boldsymbol{e}$, hence you can obtain the unique solution if the target point, $\boldsymbol{Y}_u^{k,\tau-1}$ is fixed. Let the solution of $\boldsymbol{\theta}^k$ be denoted by $\hat{\boldsymbol{\theta}}_u^{k,\tau-1}$. A set of the solutions, $\{\hat{\boldsymbol{\theta}}_u^{k,\tau-1}|u = 1, 2, \cdots, U\}$ is obtained from the set of the targets, $\mathcal{Y}^{k,\tau-1}$. The resultant set of the solutions, $\{\hat{\boldsymbol{\theta}}_u^{k,\tau-1}|u = 1, 2, \cdots, U\}$, well approximate the marginal posterior distribution, $p_k^\tau(\boldsymbol{\theta}^k|I)$, non-parametrically.

2.3.2 Inference on Graphical Model for All Segments

The conditional dependencies between the segments are then considered for updating the marginal distributions, $p_k^\tau(\boldsymbol{\theta}^k|I)$. Let $\mathcal{E}_\Lambda$ denote a set of edges of the graphical model, of which structure is determined by $\hat{\Lambda}$ in (7). Then the simultaneous posterior distribution of $\boldsymbol{\theta}^k$ is represented as follows:

$$p^\tau(\{\boldsymbol{\theta}^k\}|I) \propto \prod_k p_k^\tau(\boldsymbol{\theta}^k|I) \prod_{(k_1,k_2)\in\mathcal{E}_\Lambda} p_{k_1,k_2}(\boldsymbol{\theta}^{k1}, \boldsymbol{\theta}^{k2}), \tag{18}$$

where $p_{k_1,k_2}(\cdot,\cdot)$ is obtained by marginalizing $p(\boldsymbol{\Theta})$ in (8) with respect to all $\boldsymbol{\theta}^k$ except $\boldsymbol{\theta}^{k_1}$ and $\boldsymbol{\theta}^{k_2}$. The simultaneous distribution, $p^\tau(\{\boldsymbol{\theta}^k\}|I)$, can also be represented by a pairwise Markov random field, and we estimate the distribution of $p^\tau(\boldsymbol{\theta}^k|I)$ by inferring it on the graphical model by means of the sampling technique, for example, MCMC. The final output, $p^{\mathcal{T}}(\boldsymbol{x}_i|I)$, is obtained by marginalizing $p^{\mathcal{T}}(\boldsymbol{X}^k|I)$ with respect to all variables, $\boldsymbol{x}_{i'} \in \mathcal{P}^k$ other than $\boldsymbol{x}_i$. $p^{\mathcal{T}}(\boldsymbol{X}^k|I)$ is obtained from $p^\tau(\{\boldsymbol{\theta}^k\}|I)$ in (18) by using (16).

2.4. Experimental Method

We selected the liver as the target organ. For the experiments, we used 36 CT images, of which size was $512 \times 512 \times 512$. Some of the livers in those images have some tumors near the surfaces. The voxel size was $0.625 \times 0.625 \times 1[\text{mm}^3]$. The location, orientation and the shape of the bodies in the images were normalized based on a set of landmarks detected on the costal bones. The

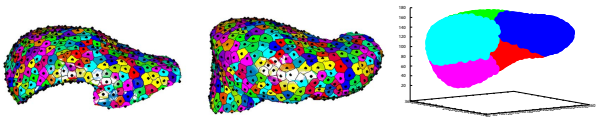

Figure 2. Examples of the corresponding points generated on two different livers, and an example of the segments ($K = 5$).

boundary of the liver in each of the images was manually labeled by an expert for constructing the SSM and for the performance evaluation.

We applied two-fold cross validation for the evaluation: The SSM was constructed from $M = 18$ surfaces and it was registered to the other 18 images for the evaluation of the accuracy of the registration. The liver surface was represented by a set of $N = 1498$ points. For constructing the SSM, we generated a set of $N = 1498$ corresponding points on each of the surfaces by means of the generalized multi-dimensional scaling (GMDS)[1], which determines the locations of corresponding points based on the geodesic distances between the points. Figure 2 shows examples of the generated corresponding points and the segments partioned by K-means method. The local appearance, I_i and $I_{ij,t}$ were determined in a cube whose side length $L = 7$, and the means and the variances of $p(I_i|\boldsymbol{x}_i)$ and $p(I_{ij,t}|\boldsymbol{x}_i, \boldsymbol{x}_j)$ were estimated based on the training images.

Then, concatenating the vectors in the k-th segment, $\{\boldsymbol{x}_i^l | i \in \mathcal{L}^k\}$, we constructed a training set of $M = 18$ samples of $\boldsymbol{X}^k$. Let the training sample of $\boldsymbol{X}^k$ obtained from the l-th surface, S^l, be denoted by $\boldsymbol{X}^{k,l}$. We estimated $\bar{\boldsymbol{X}}^k$, W, and $(\sigma^k)^2$ in (5) for each segment by applying PPCA to the training set, $\{\boldsymbol{X}^{k,l} | l = 1, 2, \cdots, 18\}$. The resultant SSM represents the location, orientation, and the shape of the segment. The global SSM that represents the conditional dependencies between the shapes of the segments was constructed from the set, $\{\boldsymbol{\Theta}^l | l = 1, 2, \cdots, 18\}$, where $\boldsymbol{\Theta}^l = [(\boldsymbol{\theta}^{1,l})^T, (\boldsymbol{\theta}^{2,l})^T, \cdots, (\boldsymbol{\theta}^{K,l})^T]^T$. Here, $\boldsymbol{\theta}^{k,l}$ was the ML estimate of the shape parameter that minimizes $\Delta'(\boldsymbol{\theta}^k) = \|\boldsymbol{X}^{k,l} - W^k\boldsymbol{\theta}^k - \bar{\boldsymbol{X}}^k\|$. The precision matrix, $\hat{\Lambda}$, was obtained by the graphical lasso, and $\boldsymbol{\Theta}$ was represented by a Gaussian: $\boldsymbol{\Theta} \propto \exp(-\boldsymbol{\Theta}^T\hat{\Lambda}\boldsymbol{\Theta})$.

We firstly evaluated the generalization ability and the specificity of the constructed SSM. Let a $3N$-vector, $\boldsymbol{X} = [(\boldsymbol{X}^1)^T, (\boldsymbol{X}^2)^T, \cdots, (\boldsymbol{X}^K)^T]^T$, represent the locations of all model points. The generalization ability of a model, G, is defined as follows[4]:

$$G = \frac{1}{n_g}\sum_{i=1}^{n_g} |\boldsymbol{X}_i' - \boldsymbol{X}_i|^2, \tag{19}$$

where $\boldsymbol{X}_i'$ is the model reconstruction of shape $\boldsymbol{X}_i$ using the model built excluding $\boldsymbol{X}_i$. Smaller value of G indicates higher generalization ability. In the experiments, $n_g = 18$, and $\boldsymbol{X}_i$ were the $n_g = 18$ data that were not used for constructing the SSM. The specificity, S, is defined as follows:

$$S = \frac{1}{n_s}\sum_{j=1}^{n_s} |\boldsymbol{X}_j - \boldsymbol{X}_j'|^2, \tag{20}$$

where $\boldsymbol{X}_j$ are shape examples generated by the model and $\boldsymbol{X}_j'$ is the nearest member of the training set to $\boldsymbol{X}_j$. Smaller value of S indicates higher specificity. We evaluated the change of G and S with respect to the number of the segments, K. While evaluating these criteria, the values of the other parameters, e.g. λ in (7) and ρ in (17), were fixed.

Then, the registration accuracy was evaluated. Let the expectation of $\boldsymbol{x}_i$ computed from the resultant posterior, $p^\tau(\boldsymbol{x}_i|I^j)$ be denoted by $\hat{\boldsymbol{x}}_i^j$. Let d_i^j denote the distance between $\hat{\boldsymbol{x}}_i$ and its nearest point on the surface, S^j, which was labeled manually for the gold standard. The registration error was evaluated with

$$E = \frac{1}{M_t}\sum_{j=1}^{M_t} E^j, \text{and } E^j = \frac{1}{N}\sum_i d_i^j, \tag{21}$$

where $M_t = 18$ is the number of images used for the evaluation. We evaluated the change of E with respect to the number of the segments, K. In addition, we compared the performances between the proposed method and the method[11] that represents the SSM with a non-hierarchical graphical model of which structure is estimated by the graphical lasso.

In the method, not only the local appearance around each point but also the appearance between two neighboring points are used for computing the likelihood, as shown in (10). The effect of $p(I_{ij}|\boldsymbol{x}_i, \boldsymbol{x}_j)$ in (10) on the registration performance was evaluated by exchanging the likelihood function, L, with the following one:

$$L'^k(\boldsymbol{X}^k) = \prod_{i \in \mathcal{L}^k} p(I_i|\boldsymbol{x}_i). \tag{22}$$

The effect of the sparsity regularization used in the registration of each segment was also evaluated by exchanging the cost function $E(\boldsymbol{\theta}^k, \boldsymbol{e})$ in (17) with

$$E'(\boldsymbol{\theta}^k) = \|\boldsymbol{Y}_u^{k,\tau^1} - W^k\boldsymbol{\theta}^k - \bar{\boldsymbol{X}}^k\|. \tag{23}$$

In the method, the structure of the graphical model of $\boldsymbol{\theta}^k$ is estimated by using the graphical lasso. The effect of this structure estimation was evaluated by using two other registration results: One was obtained when no edges existed in the graph. In this case, each segment was registered independently, and no relationship between segments was used. The other was obtained when the empirical covariance, Σ_Θ, was used for representing the relationships between the shape parameters. In this case, the graphical model was represented by a complete graph.

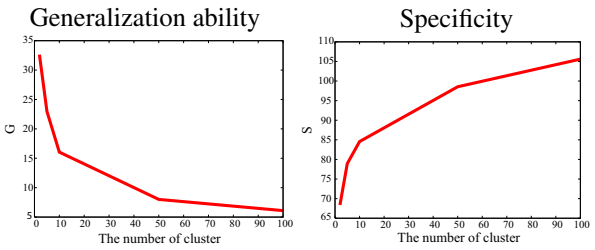

Figure 3. Change of the generalization ability (G) and of the specificity (S) with respect to K

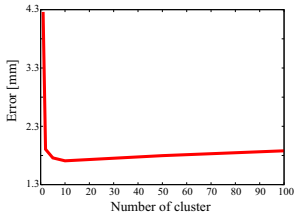

Figure 4. The change of the registration error with respect to the number of segments, K

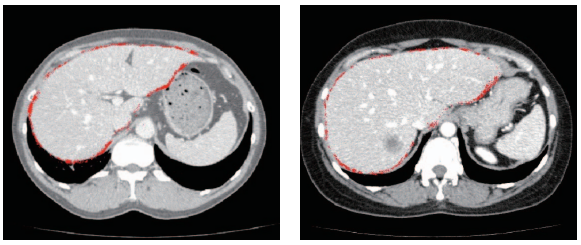

Figure 5. Two examples of the registration results. The red particles represent $p^\tau(\boldsymbol{x}_i|I)$, non-parametrically.

3. Experimental Results

3.1. Generalization Ability and Specificity

The graphs in Fig.3 show the change of the generalization ability and the specificity with respect to the number of the segments, K. In the graphs, x-axis indicates the number of K, and the y-axis indicates the generalization ability (left) and the speficity (right). As shown in the graphs, the generalization ability became higher (the value of G decreased) and the specificity became poorer (the value of S increased) monotonically with respect to the increase of K. This results show that the generalization ability was improved by making the SSM hierarchical. This improvement is important especially when the number of training data is small.

The proposed method was applied to the test images. The marginal posterior probability distribution, $p^\tau(\boldsymbol{x}_i|I)$ was computed for each point in each image. Figure 5 shows some examples of the registration results. Those figures show that the values of $p^\tau(\boldsymbol{x}_i|I)$ were high along the true surfaces, successfully. We evaluated the registration error, E, in (21), for each K. Figure 4 shows the results. The error was minimum when $K = 10$ in the experiments, hence we fixed $K = 10$ in the following experiments. It should be noted that the proposed SSM becomes a non-hierarchical PPCA-based model when $K = 1$ and that, in case $K = N$, the proposed SSM can be represented by a non-hierarchical sparse graphical model proposed in [11]. Table 1 shows the comparison of the errors, E, between the proposed method and the sparse graphical model based method proposed in [11]. In the implementation of the latter method, the value of the coefficient of the sparsity regularization, which corresponds to λ in (7), was determined so that the registration error became minimum. In the table, the minimum value, average, and the maximum value of $\{E^j\}$ are shown. The proposed method outperformed the non-hierarchical graphical model based approach, as shown at the row (a) in the table. Making the SSM hierarchical, we improved the registration performance.

Table 1. Registration error of each method

Error [mm]	min	average	max
(a) Proposed method	1.08	1.75	3.31
(b) Non-hierarchical model[11]	1.49	2.16	3.68
(c) L' (local appearance)	2.64	4.46	6.67
(d) E' (non-robust)	1.16	2.75	5.58
(e) No edges	1.09	1.84	3.33
(f) Complete graph (Σ_Θ)	1.08	1.80	3.32

Two likelihood functions, L in (10) and L' in (22) were compared based on the registration accuracy. The row (a) in Table1 shows the result obtained when L was used. The row (c) corresponds to the results obtained when L' was used. As shown, the performance was largely improved by referring to I_{ij} in the computation of the likelihood. Figure 6 demonstrates the effect of the likelihood function, L. Figure 6 (A) and (B) show the original images. Figure 6 (C) shows the results obtained when L' was used. Some model points were located on false boundaries of neighboring organs. On the other hand, the results were improved when L was used for the likelihood function, as shown in (D) in the figure.

The effect of the robust registration was also evaluated. The row (d) in Table 1 shows the results obtained without using the sparsity regularization for the segment registration. Comparing the rows (a) and (d), we found that we could improve the registration accuracy by using the sparsity regularization for the segment registration. The rows (e) and (f) in Table 1 show the results obtained when the graphical model of $\{\boldsymbol{\theta}^k\}$ had no edges (e) and when the

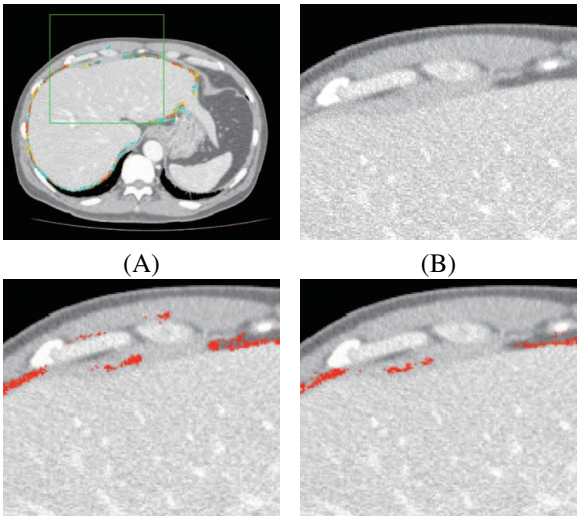

Figure 6. Comparison of registration results with respect to the difference of the likelihood. (A) and (B) show the original images. (C) and (D) were obtained when L' and L were used, respectively

graphical model was complete graph (f). As shown in the table, the registration results were the most accurate when the structure of the graphical model of $\boldsymbol{\theta}^k$ was estimated by the graphical lasso (a).

4. Conclusion

In this paper, we proposed a new non-rigid robust registration method that registers a PDM to given 3D images. The contributions of the paper were (1) the hierarchical SSM that has better generalization ability was introduced, (2) the registration algorithm of the hierarchical SSM that can estimate the marginal posterior distribution was proposed, and (3) the registration performance was improved by (3-1) robustly registering each segment with the sparsity regularization and by (3-2) referring to the appearance between the neighboring points in the likelihood computation. The SSM of a liver was constructed from a set of clinical CT images, and the performance of the proposed method was evaluated. Experimental results demonstrated that the performance was improved by using the proposed likelihood function, L, and by using the robust registration technique that uses the sparsity regularization.

Future works include to develop some efficient methods for determining the values of the parameters, λ in (7), ρ in (17), and the number of segments, K. In the experiments, we determined those values based on the registration performance evaluated with the cross-validation. This scheme is, currently, time-consuming.

References

[1] A. Bronstein, M. Bronstein, and R. Kimmel. Generalized multidimensional scaling: a framework for isometry-invariant partial surface matching. *Proceedings of the National Academy of Sciences of the United States of America*, 103(5):1168, 2006. 6

[2] T. Cootes, C. Taylor, D. Cooper, J. Graham, et al. Active shape models-their training and application. *Computer vision and image understanding*, 61(1):38–59, 1995. 1

[3] C. Davatzikos, X. Tao, and D. Shen. Hierarchical active shape models, using the wavelet transform. *Medical Imaging, IEEE Transactions on*, 22(3):414–423, 2003. 2

[4] R. Davies, C. Twining, T. Cootes, and C. Taylor. Building 3-d statistical shape models by direct optimization. *Medical Imaging, IEEE Transactions on*, 29(4):961–981, 2010. 1, 6

[5] R. Donner, B. Micušık, G. Langs, and H. Bischof. Sparse mrf appearance models for fast anatomical structure localisation. *BMVC*, 2007. 3

[6] J. Friedman, T. Hastie, and R. Tibshirani. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441, 2008. 2, 4

[7] S. Granger and X. Pennec. Multi-scale em-icp: A fast and robust approach for surface registration. *Computer Vision–ECCV 2002*, pages 69–73, 2006. 2

[8] T. Heimann and H. Meinzer. Statistical shape models for 3d medical image segmentation: A review. *Medical image analysis*, 13(4):543–563, 2009. 1

[9] J. Hermans, D. Smeets, D. Vandermeulen, and P. Suetens. Robust point set registration using em-icp with information-theoretically optimal outlier handling. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 2465–2472. IEEE, 2011. 2

[10] H. Hontani, T. Matsuno, and Y. Sawada. Robust nonrigid icp using outlier-sparsity regularization. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 174–181. IEEE, 2012. 2, 5

[11] Y. Sawada and H. Hontani. A study on graphical model structure for representing statistical shape model of point distribution model. *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2012*, pages 470–477, 2012. 2, 6, 7

[12] S. Soatto. Actionable information in vision. In *Computer Vision, 2009 IEEE 12th International Conference on*, pages 2138–2145. IEEE, 2009. 1

[13] M. Tipping and C. Bishop. Probabilistic principal component analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61(3):611–622, 1999. 3

[14] P. Tresadern, H. Bhaskar, S. Adeshina, C. Taylor, and T. Cootes. Combining local and global shape models for deformable object matching. In *Proceeding of the BMVC*, 2009. 2

[15] S. Zhang, Y. Zhan, M. Dewan, J. Huang, D. Metaxas, and X. Zhou. Towards robust and effective shape modeling: Sparse shape composition. *Medical image analysis*, 16(1):265–277, 2012. 2, 5