Sparse Gaussian graphical model of spatial distribution of anatomical landmarks – whole torso model building with training datasets of partial imaging ranges

Shouhei Hanaoka[1], Yoshitaka Masutani[1,2], Mitsutaka Nemoto[1], Yukihiro Nomura[1], Soichiro Miki[3], Takeharu Yoshikawa[3], Naoto Hayashi[3], Kuni Ohtomo[1,2]

[1] Department of Radiology, [2] Division of Radiology and Biomedical Engineering, Graduate School of Medicine and [3] Department of Computational Diagnostic Radiology and Preventive Medicine, The University of Tokyo Hospital, 7-3-1 Hongo, Bunkyo-ku, Tokyo, Japan

hanaoka-tky@umin.ac.jp

Abstract. A method of building a point distribution model for multiple anatomical landmarks from training datasets with various different imaging ranges is presented. To cope with the missing data problem caused by partial imaging ranges of training datasets, the MissGLasso algorithm is applied to build the model. The joint probability distribution of logarithmic distances between all landmark pairs is approximated as a sparse Gaussian graphical model which has a sparse precision matrix calculated by the graphical lasso method. Additionally, EM algorithm is also utilized to apply graphical lasso to training sample vectors which include many missing values. To evaluate models built by the proposed method, the models were embedded into multiple-landmark detection systems and the detection sensitivities were compared between models trained with partial and entire imaging range datasets. The overall sensitivities of landmark detection using these two models were 72.78 and 75.02%, respectively. Because the detection sensitivity for most of the landmarks differs little between the two models, it can be concluded that the MissGLasso algorithm could effectively handle missing values in training landmark distribution models.

Keywords: Sparse Gaussian graphical model, Landmark, Graphical lasso

1 Introduction

The automatic detection of anatomical landmark positions often plays a key role in various medical image analyses, such as organ segmentation, interindividual or intermodality image registration or computer-aided lesion detection. Various point distribution models (PDMs) of the spatial distribution of multiple anatomical landmarks have been used to detect a series of landmarks [1][2][3]. For example, in [1] Seifert et al. used predefined spatial constraints between specific pairs of landmarks in detecting 19 body trunk landmarks. In [2], Potesil et al. proposed the pictorial structure

model in which a graph structure connects 22 landmarks from cervical to pelvic structures. Although the graph structure itself is arbitrarily predefined, the spatial constraint between two connected landmarks is learned from training datasets. On the other hand, in [3] Hanaoka et al. built a statistical PDM on 173 landmark positions and applied it to a landmark detection task. Their target imaging range was neck-to-pelvis and the head was not included. In all these methods, PDMs play an important role by providing prior knowledge on the human body structure. Therefore, it is suggested that the quality of PDM will strongly affect the overall performance of landmark detection applications.

Theoretically, a landmark-based PDM for the whole human body can be built in the same manner as the methods mentioned above. Such a whole body PDM will be more versatile than partial-body PDMs, that is, it can be applied to a wider range of medical image processing and to various organs. However, to the best of our knowledge, no research on building a PDM for the whole body by means of statistical modeling has been reported. It is considered that the main reason for this is the difficulty of preparing a sufficient amount of whole-body training datasets. Medical image examinations, such as computed tomography (CT) or magnetic resonance imaging (MRI), are rarely performed for the whole body because of limited imaging machine resources or excess radiation exposure. If only partial-body datasets with various imaging ranges (e.g. chest only, abdomen only, etc.) are available as training datasets, the usual statistical model estimation methods will not be applicable. Instead, an alternative method that can handle missing values in the training datasets will be required. One of the aims of this study is to overcome this problem by applying an EM-algorithm-based missing value imputation method [4] in estimating the statistical landmark distribution model.

The other aim of this study is to apply the sparse Gaussian graphical model [5] to landmark distribution modeling. In particular, we focused on estimating the precision matrix (inverse covariance matrix) on the interlandmark distances between landmark pairs as a sparse matrix. We chose the interlandmark distances as the model variables, instead of the landmark coordinates themselves, because of their desirable features such as spatial rotation invariance and robustness to local deformations. Generally speaking, it can be expected that a certain distribution can be statistically modeled better with a sparse precision matrix if most of the variable pairs are conditionally independent. This is not the case for the interlandmark distances, however, because most of the distances have strong positive correlations with each other owing to the scale factor. A large person has larger interlandmark distances, and vice versa. Nevertheless, we found that the distances become much less correlated after the scale factor is normalized appropriately. Thus, we assumed that the precision matrix can be estimated as a sparse matrix if the effect of the scale factor is removed in advance. On the basis of this assumption, we developed a novel method to estimate the sparse precision matrix while handling the scale factor separately.

For these two purposes, we propose a method for building a sparse Gaussian graphical model of interlandmark distances from training datasets with insufficient imaging ranges. The method is based on the MissGLasso method, proposed by Städler et al. [4], which consists of two algorithms: (1) an expectation-maximization

(EM) algorithm to handle missing data and (2) the graphical lasso method [5] for estimating the precision matrix with l_1 norm regularization. We modified the original MissGLasso method so that both the scale factor and the scale-normalized interlandmark logarithmic distance set can be modeled simultaneously. The proposed method was evaluated with 78 chest-to-pelvis CT volume datasets. For comparison, two models were built: (1) a statistical landmark distribution model built with intentionally cropped training volumes and (2) a model built with the whole imaging ranges. Then, the landmark detection performances using these models were compared to evaluate the applicability of the proposed method to partial volume training datasets.

2 Methods

2.1 Landmark distribution model based on interlandmark distances

In this study, the spatial landmark distribution is modeled as a joint probability function whose variables are the interlandmark logarithmic distances between all landmark pairs. Here, the reason why logarithmic distances are used is that it enables the proposed algorithm to handle the scale factor as a linear factor. Additionally, since the domain of log-distances is not restricted to positive but can have negative values, they may be more suitable to be modeled with Gaussian distribution.

Suppose that the number of landmarks is L and that their coordinates are $\mathbf{x}_i,\ i \in \{1,2,3,\dots,L\}$. Then, the logarithmic distances are defined as $d_{i,j} = \log|\mathbf{x}_i - \mathbf{x}_j|, 1 \leq i < j \leq L$. Let the concatenated vector of all $d_{i,j}$ be $\mathbf{d} = \left(d_{1,1}d_{1,2}\cdots d_{i,j}\cdots d_{L-1,L}\right)^t$. Suppose that the mean vector of $\mathbf{d}$ is estimated as $\bar{\mathbf{d}}$ and the precision matrix is estimated as $\mathbf{K}$, then the probability function $p(\mathbf{X})$ for any landmark position set $\mathbf{X} = \{\mathbf{x}_i\}$ can be estimated as a multidimensional Gaussian distribution as follows:

$$p(\mathbf{X}) = \sqrt{2\pi}^{-M} \cdot \sqrt{|\mathbf{K}|} \cdot \exp\left\{-\frac{1}{2}\left(\mathbf{d}-\bar{\mathbf{d}}\right)^t \mathbf{K}\left(\mathbf{d}-\bar{\mathbf{d}}\right)\right\}. \quad (1)$$

where $M = {}_L\mathrm{C}_2$ is the dimension number of $\mathbf{d}$. Therefore, only the mean vector $\bar{\mathbf{d}}$ and precision matrix $\mathbf{K}$ must be determined to define a landmark distribution model. $\bar{\mathbf{d}}$ and $\mathbf{K}$ can be estimated from manually input ground-truth landmark positions in the training datasets. The main topic of this study is how to adequately estimate $\bar{\mathbf{d}}$ and $\mathbf{K}$ where the number of training datasets N is far less than the model dimension ${}_L\mathrm{C}_2$ and many missing values are included.

Note that the distribution (1) is translation- and rotation-invariant because only the interlandmark distances are considered. Moreover, the logarithmic distance vector $\mathbf{d}$ is altered linearly by a scale transformation; that is, when a scale transformation with scale factor α is applied to the landmark position set (i.e., $\mathbf{x}_i \mapsto \alpha \cdot \mathbf{x}_i, \forall \mathrm{i}$), the vector $\mathbf{d}$ will be translated as $\mathbf{d} \mapsto \mathbf{d} + \log\alpha \cdot \mathbf{1}$. This is a simple translation parallel to the vector $\mathbf{1} = (1\ 1\ \dots\ 1)^{\mathrm{t}}$.

2.2 Graphical lasso with scale factor normalization

Graphical lasso [5] is a method of estimating a sparse precision matrix $\mathbf{\Theta}$ from a given sample covariance matrix $\mathbf{S}$ by maximizing the l_1 norm-penalized log likelihood as follows:

$$\log\det\mathbf{\Theta} - \mathrm{trace}(\mathbf{S\Theta}) - \rho\|\mathbf{\Theta}\|_1. \tag{2}$$

Here, ρ is a parameter that controls the sparsity of $\mathbf{\Theta}$.

The sample covariance matrix $\mathbf{S}$ can be calculated from the set of the vector $\mathbf{d}$ in the given training datasets. However, as described above, $\mathbf{S}$ is not suitable for the graphical lasso estimation because of the scale factor. To overcome this problem, $\mathbf{S}$ is divided into two: the covariance matrix of the scale-normalized log-distances $\mathbf{S}_{\mathrm{nor}}$ and the variance of the scale factor γ. After this division, $\mathbf{S}_{\mathrm{nor}}$ is processed by the graphical lasso method to estimate the precision matrix.

The division of $\mathbf{S}$ is performed as follows. Firstly, the scale factor of a given vector $\mathbf{d}$ is defined as $\mathrm{scale}(\mathbf{d}) = \left(\mathbf{d} - \bar{\mathbf{d}}\right) \cdot \left(\frac{1}{M} \cdot \mathbf{1}\right)$. Here, $\bar{\mathbf{d}}$ is the sample mean of $\mathbf{d}$ and $\mathrm{scale}(\mathbf{d})$ is the average of M elements of the mean-subtracted vector $\left(\mathbf{d} - \bar{\mathbf{d}}\right)$. The variance of the scale factors is calculated as the variance of all $\mathrm{scale}(\mathbf{d})$ from the datasets. Secondly, the scale-normalized log-distance is defined as $\mathbf{d}_{\mathrm{nor}} = \mathbf{d} - \bar{\mathbf{d}} - \mathrm{scale}(\mathbf{d}) \cdot \left(\frac{1}{M} \cdot \mathbf{1}\right)$. Then the covariance matrix $\mathbf{S}_{\mathrm{nor}}$ is calculated from all $\mathbf{d}_{\mathrm{nor}}$ extracted from training datasets.

The precision matrix of $\mathbf{d}_{\mathrm{nor}}$, namely $\mathbf{\Theta}_{\mathrm{nor}}$, is estimated using the graphical lasso algorithm by maximizing (2) with the given $\mathbf{S}_{\mathrm{nor}}$. Note that $\mathbf{K}$ (the precision matrix of $\mathbf{d}$) can be estimated from $\mathbf{\Theta}_{\mathrm{nor}}$ as follows: Assuming that the scale factor $\mathrm{scale}(\mathbf{d})$ and the normalized distance vector $\mathbf{d}_{\mathrm{nor}}$ are statistically independent, a simple addition rule of covariance matrices $\mathbf{K}^{-1} = \mathbf{\Theta}_{\mathrm{nor}}^{-1} + \gamma \cdot \mathbf{1}\mathbf{1}^t$ is satisfied. ($\gamma \cdot \mathbf{1}\mathbf{1}^t$ corresponds to the covariance matrix of the scale factor.) Then, from the Sherman-Morrison formula,

$$\mathbf{K} = (\mathbf{\Theta}_{\mathrm{nor}}^{-1} + \gamma \cdot \mathbf{1}\mathbf{1}^t)^{-1} = \mathbf{\Theta}_{\mathrm{nor}} - \gamma/(1 + \gamma \cdot \mathbf{1}^t\mathbf{\Theta}_{\mathrm{nor}}\mathbf{1}) \cdot \mathbf{\Theta}_{\mathrm{nor}}\mathbf{1}\mathbf{1}^t\mathbf{\Theta}_{\mathrm{nor}} \tag{3}$$

is satisfied. Note that using this equation, the probability calculation in Eq. (1) can be speeded up by taking advantage of high sparsity of the matrix $\mathbf{\Theta}_{\mathrm{nor}}$.

A modification was introduced to the original graphical lasso method in this study. It is to change the l_1-norm-penalizing weights for each variable according to its standard deviation. In graphical lasso it is possible to change the sparsity control factor ρ for each element of $\mathbf{\Theta}$ [5]. This is performed by modifying Eq. (2) as follows:

$$\log\det\mathbf{\Theta} - \mathrm{trace}(\mathbf{S\Theta}) - \|\mathbf{\Theta} * \mathbf{P}\|_1 \tag{4}$$

where $\mathbf{P} = \{\rho_{jk}\}$ with $\rho_{jk} = \rho_{kj}$, and * indicates component-wise multiplication. On this basis, we determined the element-wise l_1-norm-penalizing weight as follows:

$$\rho_{jk} = \rho \cdot f\left(\sigma_j\right) \cdot f(\sigma_k) \tag{5}$$

$$f(x) = \begin{cases} x & \text{if } \sigma_{threshold} \leq x \\ \sigma_{threshold} & \text{if } x < \sigma_{threshold} \end{cases}$$

where σ_i is the standard deviation of the i-th element of $\mathbf{d}$, and $\sigma_{threshold}$ is a parameter to be determined in advance. The motivation for introducing this weight is to avoid the excessive effect of variables with very large variances. We found that this weight modification is very useful for improving the result.

2.3 EM algorithm and MissGLasso

Most of the methods of estimating the distribution from given sample data require that no values are missing from the given sample. However, this is not the case when the imaging range is limited and some landmarks are out of range. One element in $\mathbf{d}$ (i.e., one interlandmark distance) will be "missed" if either of the two landmarks is out of the imaging range. Therefore, a significant number of elements in $\mathbf{d}$ from a partial volume training dataset may be missed. One frequently used solution for such a "missing value" problem is to apply the EM algorithm. In this study we utilized the MissGLasso method, which is a fusion of the EM algorithm and the graphical lasso method.

Details of the MissGLasso method are available in [4]. In brief, the model parameter $\mathbf{\Theta}_{\text{nor}}$ is iteratively updated in MissGLasso by alternately applying E and M steps. (Fig. 1, left) In the E step, the covariance matrix and mean vector are estimated from both the observed (nonmissed) values and recently estimated $\mathbf{\Theta}_{\text{nor}}$. Using the updated covariance matrix and mean vector, the precision matrix $\mathbf{\Theta}_{\text{nor}}$ is estimated and updated in the following M step. In MissGLasso, the M step is virtually the graphical lasso method itself. In this study, we iterated the EM algorithm 10 times to find an optimal $\mathbf{\Theta}_{\text{nor}}$ from partial volume training datasets.

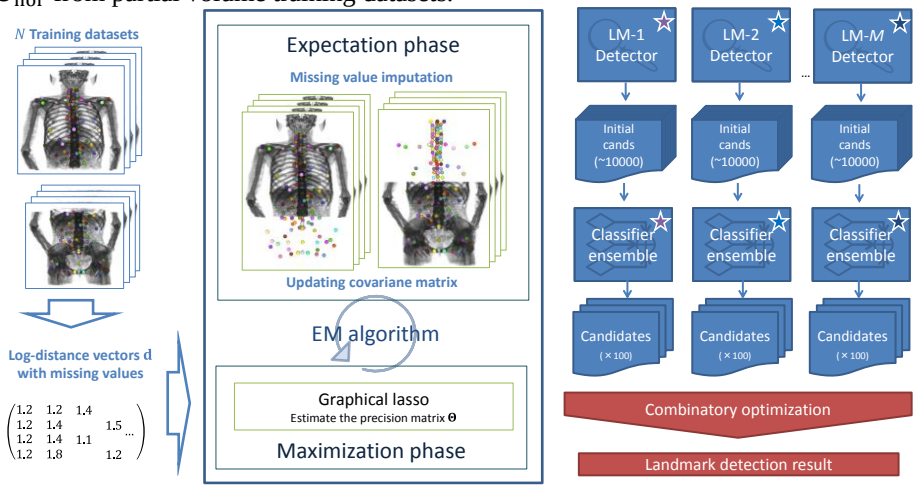

Fig. 1. Outline of proposed method. (Left) MissGLasso-based model estimation. (Right) Landmark detection system used in evaluation.

2.4 Evaluation

The positions of a total of 138 (64 chest, 37 abdominal, 37 pelvic) landmarks in the human torso were modeled and detected in this study. We compared the landmark detection performances using two different models. The *GLasso* model was trained with ground truth landmark positions from full-range (chest-to-pelvis) CT volumes. On the other hand, the *MissGLasso* model was trained with not all the ground truth positions. Instead, the pelvic landmarks were removed from half of the training datasets and the chest landmarks were removed from the other half. Thus, these two halves simulate chest-to-abdomen and abdomen-to-pelvis imaging ranges, respectively.

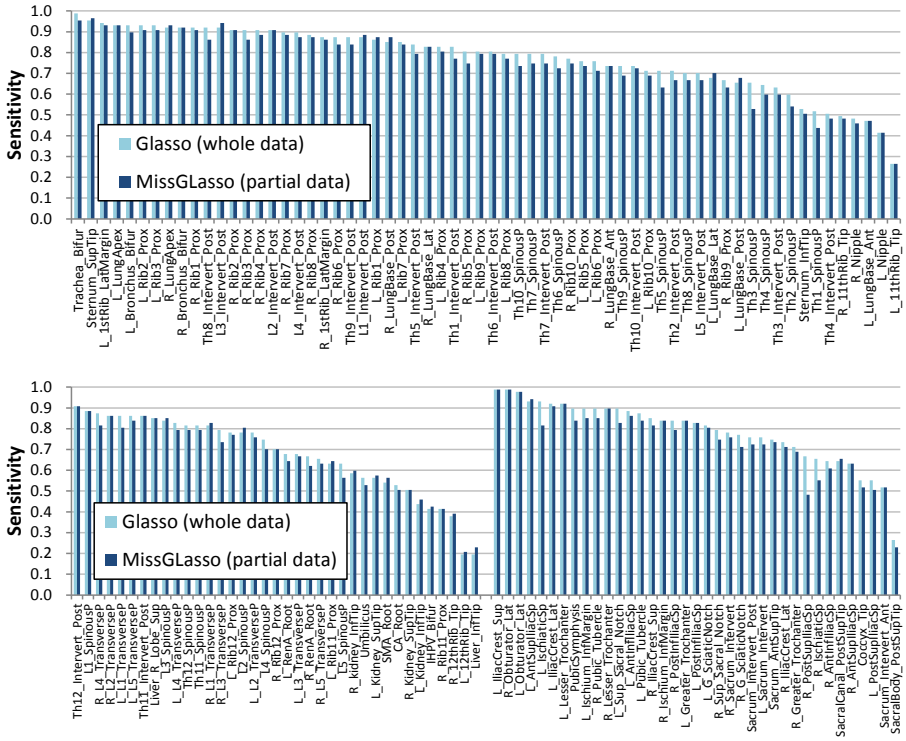

Fig. 2. Result of landmark detection performance. (Top) Sensitivities for thoracic landmarks. (Bottom, from the left) Sensitivities for abdominal and pelvic landmarks. All landmarks are sorted by the sensitivities using the *GLasso* model.

A total 87 of CT datasets of healthy or diseased human torsos without an intravenous contrast agent were used in the evaluation. The model training was performed with 4-fold cross-validation; 87 datasets were divided into 4 groups, and the datasets in each group were tested with a model that was trained using the other 3 groups. For *MissGLasso* models, these training datasets were merged and divided again into chest-to-abdomen and abdomen-to-pelvis imaging range simulation subgroups. The values of ρ and $\sigma_{\text{threshold}}$ were determined empirically as 0.75 and 0.2, respectively.

To evaluate the suitability of the estimated models, a multiple-landmark detection method proposed in [3] (Fig. 1, right) was applied to the test datasets and the detection performances using *GLasso* and *MissGLasso* models were compared. The outline of the landmark detection method used is as follows. Firstly, each of the target landmarks is detected independently by a landmark-specific detector that is composed of an appearance-model-based initial detector and a classifier ensemble. Each detector searches possible landmark positions within the given CT volume and then outputs a number of candidate points (100 candidates in this study). Then, the following combinatory optimization algorithm chooses the most suitable combination of all landmark positions from the lists of candidates. In the latter phase, a landmark distribution model [Eq. (1)] is used as prior knowledge to seek the best combination. The optimization problem is formulated by maximum *a posteriori* (MAP) estimation and is solved by Gibbs sampling and simulated annealing methods. Each detection result was evaluated as successful if the detected point was within 2 cm from the manually input ground truth point.

Both the model estimation and landmark detection system were implemented on and experimented with a computer with Intel Xeon E5-2640 CPU×2, 64 GB RAM and nVidia Tesla K20m GPGPU. It took approximately 2 h to estimate one *MissGLasso* model and 30 min for one landmark detection task.

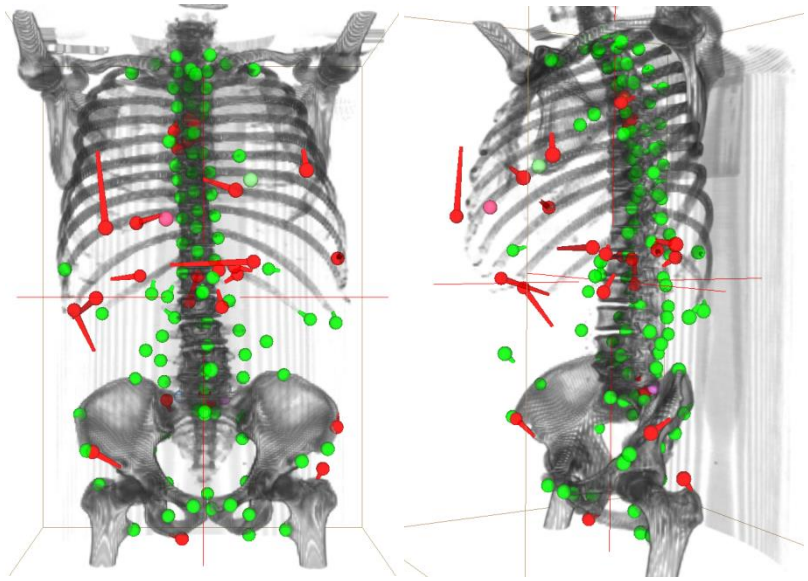

Fig. 3. An example of landmark detection result. (Left) a frontal view. (Right) a left anterior oblique view. Green and red pins represent successfully and wrongly detected landmarks, respectively. The pin heads represent the detected positions, whereas the pin tails are the true landmark positions. Most of wrongly detected landmarks were those defined on soft-tissue structures.

3 Results

The overall result is shown in Fig. 2. The averages and standard deviations of sensitivities were 75.02% ± 0.17 and 72.78% ± 0.17 with *GLasso* and *MissGLasso*, respectively. Within the detection failures of both results (24.98% and 27.22%, respectively), 9.92% were due to the detection phase (i.e., no true position was outputted by the detector as a candidate). Thus, the failures due to the combinatory optimization phase using the models were roughly estimated as 15.06% and 17.30%, respectively.

Because the detection sensitivities for most of the landmarks differ little between *GLasso* and *MissGLasso*, it can be concluded that the MissGLasso algorithm could effectively handle missing values in the training datasets. The detection performance was similar even for landmarks with very poor sensitivity.

A performance comparison with a related study recently reported by Liu et al. [8] was performed on 5 landmarks (all landmarks shared by their study and ours). The mean errors of our MissGLasso setting were 4.4, 8.4, 7.8, 10.8, 40.6 (mm) for trachea bifurcation, left and right lung tops, liver top and bottom, respectively. In Liu et al., the corresponding errors were 2.5, 2.6, 3.2, 2.5 and 6.4 (mm), respectively. Though our result showed less accuracy than [8], we believe that our results for 4 landmarks, aside from the bottom of the liver, are not so poor if ambiguities of landmark positions are taken into account. Moreover, it may be improved by any appropriate post-process (e.g., [7]).

4 Discussion

A method of estimating a landmark point distribution model from training datasets with various partial imaging ranges was presented. The deterioration of the landmark detection performance was minimal using the model estimated with only chest-to-abdomen and abdomen-to-pelvis image datasets. Therefore, it was suggested that the proposed method can be a key technique for building landmark distribution models with a wider range, or ultimately, a whole body landmark PDM.

Application of our method to build a whole-body PDM is not limited to landmark detection. Possible applications include estimation or imputation of structures unseen in a given image (due to imaging range or nature of the modality). For example, the model may estimate the body height from the landmark positions of the pelvis only.

The main motivation of introducing sparsity in this study is to handle the High-Dimensional Low Sample Size (HDLSS) problem better. Regularization is a critical issue for HDLSS. L1-norm regularization is a state-of-the-art method and its superiority against classical L2-norm regularization has been suggested by many studies [9]. Our precision matrices were very sparse, having approx. 0.1% of non-zero elements.

This study has several limitations. Firstly, the proposed method was not compared with other precision matrix estimation methods such as Tikhonov regularization. However, the landmark detection system evaluates the distribution probability (Eq. (1)) about 15 million times per detection task under our current implementation and experimental conditions. The speed of this calculation completely depends upon the

sparsity of the precision matrix in this study. Thus it is expected that landmark detection with a dense precision matrix is virtually impossible. Although in [6] Hanaoka et al. proposed an effective way of calculating Eq. (1) without explicitly considering the entire precision matrix, it instead depends on explicit training sample vectors and cannot be applied to a problem with missing values.

Secondly, the detection sensitivity is not sufficient for some landmarks. For example, the sensitivities of 42 landmarks out of 158 were less than 70% with the criterion of 2 cm from the ground truth. Improvement of the detection performance is an issue in the future. Possible solutions include the landmark position fine-tuning algorithm proposed by Nemoto et al [7] or parameter tuning for each landmark detector [2].

As a future work, we are planning to apply the proposed method to building a multimodality landmark PDM. For example, suppose that you have a number of CT datasets with manually input landmark positions, and you also have other MRI datasets with landmark positions input. If the two datasets share some of the landmarks, the proposed method can build a combined landmark distribution model from both datasets. Such a combined model will be useful in various ways. For instance, using a CT-MRI combined model, the position of a certain landmark that is only visible in MRI can be estimated in any given CT images. Our final goal is to represent all landmark positions in the whole body and in any medical images in a single landmark distribution model.

Acknowledgement

This study is a part of the research project "Computational Anatomy for Computer-aided Diagnosis and therapy: Frontiers of Medical Image Sciences", supported by a grant-in-aid for scientific research on innovative areas MEXT, Japan.

References

1. Seifert, S., Barbu, A., Zhou, S. K., Liu, D., Feulner, J., Huber, M., Suehling, M., Cavallaro, A., Comaniciu, D.: Hierarchical parsing and semantic navigation of full body CT data. In Proc. of SPIE 7259 (2009)
2. Potesil V., Kadir T., Platsch G., Brady M.: Improved Anatomical Landmark Localization in Medical Images Using Dense Matching of Graphical Models. In Proc. of BMVC 1-10 (2010)
3. Hanaoka S., Masutani Y., Nemoto M., Nomura Y., Yoshikawa T., Hayashi N., Yoshioka N. and Ohtomo K.: An improved multiple anatomical landmark detection method with combinatorial optimization and Madaboost-based candidate likelihood determination. Int. J. CARS, 7 Suppl.1, S330-331 (2012)
4. Städler N, Bühlmann P.: Missing values: sparse inverse covariance estimation and an extension to sparse regression. Statistics and Computing 22, 219-235 (2012)
5. Friedman J, Hastie T, Tibshirani R: Sparse inverse covariance estimation with the graphical Lasso. Biostatistics 9, 432–441 (2008)
6. Hanaoka S., Masutani Y., Nemoto M., Nomura Y., Yoshikawa T., Hayashi N., Yoshioka N. and Ohtomo K.: Probabilistic Modeling of Landmark Distances and Structure for

Anomaly-proof Landmark Detection. In: Proceedings of the Third International Workshop on Mathematical Foundations of Computational Anatomy 159-169 (2011)
7. Nemoto M., Masutani Y., Hanaoka S., Nomura Y., Miki S., Yoshikawa T., Hayashi N. and Ohtomo K.: Coarse-to-fine localization of anatomical landmarks in CT images based on multi-scale local appearance and rotation-invariant spatial landmark distribution model. In Proc. of SPIE 8669 (2013)
8. Liu D, Zhou S. K. Anatomical Landmark Detection Using Nearest Neighbor Matching and Submodular Optimization. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2012. Lecture Notes in Computer Science Volume 7512, pp 393-401 (2012)
9. Honorio J., Samaras D.: Multi-task learning of Gaussian graphical models. Proc. of the 27th International Conference on Machine Learning, Haifa, Israel, 2010.